CHAPTER 8

USING COMPUTER-ASSISTED TEXT ANALYSIS (CATA) TO INFORM EMPLOYMENT DECISIONS: APPROACHES, SOFTWARE, AND FINDINGS

Emily D. Campion and Michael A. Campion

ABSTRACT

This literature review is on advanced computer analytics, which is a major trend in the field of Human Resource Management (HRM). The authors focus specifically on computer-assisted text analysis (CATA) because text data are a prevalent yet vastly underutilized data source in organizations. The authors gathered 341 articles that use, review, or promote CATA in the management literature. This review complements existing reviews in several ways including an emphasis on CATA in the management literature, a description of the types of software and their advantages, and a unique emphasis on findings in employment. This examination of CATA relative to employment is based on 66 studies (of the 341) that bear on measuring constructs potentially relevant to hiring decisions. The authors also briefly consider the broader machine learning literature using CATA outside management (e.g., data science) to derive relevant insights for management scholars. Finally, the authors discuss the main challenges when using CATA for employment, and provide recommendations on how to manage such challenges. In all, the authors hope to demystify and encourage the use of CATA in HRM scholarship.

Research in Personnel and Human Resources Management
Research in Personnel and Human Resources Management, Volume 38, 285–325

ISSN: 0742-7301/doi:10.1108/S0742-730120200000038010

Keywords: Computer-aided text analysis; CATA; employment decisions; selection; machine learning; literature review

Advanced computer analytics is a major trend in the field of Human Resource Management (HRM) and its primary underlying scientific discipline, Industrial and Organizational Psychology.[1] Scholars and organizations alike have become increasingly interested in using computer-assisted text analysis (CATA) to examine text data. The ability to analyze text data is a major breakthrough due to the extensive availability of text data in HRM, the potential for the improved use of that data for making HRM decisions, and the increased objectivity, efficiency, and accuracy of measurement that may result. To demystify and promote CATA, we review the CATA research in management. This review will complement existing reviews (e.g., McKenny, Aguinis, Short, & Anglin, 2018; McKenny, Short, & Payne, 2013; Short, McKenny, & Reid, 2018) in several ways: (1) we build on these reviews by including nearly an additional 200 articles (e.g., Short et al., 2018), (2) we describe the types of software for all types of CATA from the more rudimentary (e.g., Nvivo) to the more advanced (e.g., SPSS Modeler) and the advantages of each type, and (3) we present an in-depth examination of findings, challenges, and recommendations for its uses for employment decisions.

LITERATURE COLLECTION METHOD

We conducted a comprehensive literature review in the traditional manner, starting with a computer search of all related keywords and CATA software in 16 high-quality management journals and those that publish CATA research often to generate our database of management literature that uses CATA.[2] We found 585 articles across organizational behavior, human resources, strategy, and entrepreneurship. We read the titles, abstracts, and methods of each article and eliminated those that did not apply or review CATA methods, which yielded 341 articles in our final database. Of these, 324 directly measured constructs and the remaining 17 were either reviews or provided direction on how to conduct CATA.

The initial scope of the review was meant to be all-inclusive of any literature on the topic, but the relevance of the various bodies of literature narrowed our scope for three reasons. First, because this is a relatively new and rapidly evolving area of research, most of the relevant literature in management has been published recently (81.5% of the articles in our review were published between 2010 and 2019). Second, there is a vast literature on machine learning that did not include CATA, and thus was not directly relevant to the review. Third, the focus was on the literature that bears on the measurement of human attributes, which is primarily the research in management (psychology, organizational behavior, and human resources) as opposed to other disciplines (e.g., data science). We briefly scanned the text mining literature in other disciplines for lessons that might be relevant, such as methodologies, but we do not review this literature comprehensively because they either do not use text data or their findings were not directly translatable to the management literature. Fourth, there is a large literature that uses CATA to measure writing skill in education. This was the first major area of CATA literature and consists of several hundred articles. This literature will not be included because writing skill development among students is not directly useful for employment decisions.[3]

OVERVIEW OF CATA

Before describing the findings in the literature, it might be helpful to identify the various ways to analyze text, including CATA, as a backdrop. We believe the range of approaches include, but are not limited to, the following:

1. *Using traditional human judgment-based content analysis*. This is the most basic approach to text analysis. Here, human judges such as subject matter experts (SMEs) read and categorize the sample of textual data based on similarity of the content. This technique is sometimes called a "Q Sort." With this approach, the text data (e.g., responses) are ordinarily grouped into only one category. After sorting, the human judges read the data in each category and assign descriptive labels. Often, they will count the number of responses in each category as an indication of its importance. This is not computer assisted, other than perhaps to record the information. However, it provides an important perspective because it is by far the most common approach historically and human judges still play an important role in the more automated approaches below.
2. *Using basic computer automation*. This technique is largely used to identify content by simply counting the frequencies of various words and categorizing them. They may also help visually display the results such as in "word clouds." This approach can be purely empirical by only using the variables identified by the software, or the researcher can improve the variables by modifying the categories. This involves reviewing the concepts extracted and combining or separating them based on their meaning in the context of the study, just like traditional content analysis. The computer does not know that different terms or phrases may be synonyms, but most software will allow the researcher to tell that to the computer. The initial variables extracted can be modified using other analyses to reduce the data, such as factor or cluster analyses. SMEs might also be used to help impart meaning on the categories or to check the coding and make modifications. This use of CATA has been very common in the management literature, especially for inductive research (e.g., grounded theory). Of the 324 empirical articles in our review that used CATA to measure attributes, 153 (47.2%) fell into this category.

   The most common software packages that conduct simple content analyses are Nvivo and Atlas.ti. Although they help suggest potential categories by identifying the common words, they serve as more of a data management tool for qualitative (text) data and involve little automation. They are word processing tools that facilitate text analysis by allowing the researcher to easily sort, track, and code text data, but provide little assistance in identifying the underlying meaning of the data compared with more advanced methods described below.

3. *Using rationally developed data dictionaries*. This has been a popular technique to examine the sentiment of text based on existing dictionaries. However, researchers can also create their own dictionaries to measure specific constructs and use them for content analysis. This is much like a keyword search where the researcher identifies all the relevant terms regarding a construct and then searches the documents for these words. The difference between this method and a keyword search is that the words have been identified as being reflective of various sentiments based on prior research and presumably validated. Sentiments captured can be as simple as positive and negative tone or more complex and nuanced such as specific attitudes and dispositions. The frequency with which respondents use words in the dictionary usually provides the measure of the attribute. Studies using dictionaries are more deductive because researchers are either trying to measure particular attributes for which the dictionaries have been validated *a priori*, or they are developing their own dictionary based on a known construct that is also known *a priori*. Many dictionaries are available publicly to assess various sentiments or emotions (especially the Linguistic Inquiry and Word Count [LIWC] and the DICTION software packages). Using CATA to identify sentiments has been very common in the management CATA literature. More than a third of the articles in our review used dictionaries (122; 35.8%)
4. *Using more advanced text mining software to identify the constructs underlying combinations of words that exist in a corpus (e.g., set of documents)*. This often involves natural language processing (NLP) techniques (e.g., latent semantic analysis [LSA]; latent Dirichlet analysis) that include extracting the meaningfulness of narrative information by analyzing the relationships among multiple words, as described in more detail in the software review section of this report. Advanced techniques such as NLP can be used in addition to data dictionaries and basic automation, and some software programs will combine all three. Common software programs that perform all of these analyses include R, Python, SPSS, and SAS.
5. *Using text mining software to identify concepts and combinations of words based on their predictiveness of some criterion (e.g., job performance, training performance, other outcomes, etc.)*. This begins with the approaches above, but then identifies the most useful concepts and variables from the large number extracted based on their empirical relationships with criteria. The number of variables extracted depends on the size and variation of the textual data in the corpus, but can range up to the hundreds or even thousands. This approach can also be combined with training the computer, which includes combining, deleting, relabeling, identifying synonyms, and other adjustments to improve the accuracy of the model. A variant of this is to use criterion data from SMEs to text mine against. Where criterion data do not exist or have severe limitations, or where the goal is to emulate a human judge, SMEs can score a set of written text samples and their scores can be used as the criterion. If done well, the sample of text scored would not have to be excessively large (e.g., in the hundreds) because the research protocol can be structured to ensure

wide variance, reliability, and content validity. This and the previous type of CATA fall under the broader umbrella term of "machine learning." Twenty-two (6.8%) of the 324 empirical articles in our review used machine learning.

TYPES OF STUDIES, TEXT DATA, AND CATA SOFTWARE

Table 1 shows the frequency and percentage of each of the various types of distinctions between the articles in the review. In the sections below, we describe and summarize the key findings within each type.

Table 1. Frequencies of Types of Articles in the Management Literature on CATA.

Types of Studies	Number	Percentage
Qualitative and quantitative	171	50.2%
Qualitative only	153	44.9%
Review	17	5.0%
Total	341	100%
Types of Textual Data	**Number**	**Percentage**
Transcripts (interviews, speeches, phone calls)	164	50.6%
Report (annual reports, sustainability reports)	55	17.0%
Observations	51	15.7%
Archival data	48	14.8%
News-related documents (news articles, press releases)	47	14.5%
Open-ended responses	18	5.6%
Journals articles and abstracts	15	4.6%
Web-related content (webpages, online communication)	15	4.6%
Other (cases, mission statements, crowdfunding campaigns)	40	12.4%
Types of Software	**Number**	**Percentage**
Nvivo	138	42.6%
Linguistic Inquiry Word Count (LIWC)	82	25.3%
Atlas.ti	30	9.3%
DICTION	27	8.3%
R	7	2.2%
Cat Scanner	5	1.5%
MonoConc Pro	5	1.5%
Python	4	1.2%
General Inquirer (GI)	4	1.2%
SPSS Modeler	3	0.9%
VBPro	3	0.9%
Wordstat	3	0.9%
Leximancer	2	0.6%
Textpak4	2	0.6%
Textual Analysis Computing Tools (TACT)	2	0.6%
Automap	2	0.6%
Excel	2	0.6%
Used only once	25	7.7%
Did not disclose	22	6.8%

Note: Percentages reported under "Types of Study" are out of 341. Percentages reported under "Types of Textual Data" and "Types of Software" are out of 324 – the number of empirical articles in our review.

Types of Studies

Three types of studies emerged in the literature review: (1) studies using qualitative CATA methods only, (2) studies using both qualitative and quantitative methods, and (3) reviews of the literature or articles promoting CATA. The distinction between the first and the second types of studies is whether the researchers quantified the qualitative data. That is, if the data were converted into indices and metrics used for statistical analysis like correlations, then they fell into the second type of study. Of the 341 articles in this review, 153 (44.9%) were qualitative only and generally took inductive approaches using interviews with organizational informants (e.g., respondents, experts), observations, and archival organizational data to develop constructs and flesh out processes among constructs to expand or generate management theory (Edmondson & McManus, 2007). Qualitative studies can be conducted in a number of ways, but one commonly cited approach is called *grounded theory*. According to Gephart, Gibson, and Gibbs (2004, p. 459):

> *Grounded theorizing* (Glaser & Strauss, 1967) is the process of iteratively and inductively constructing theory from observations using a process of theoretical sampling in which emergent insights direct selection and inclusion of the "next" informant or slice of data. Grounded theory involves constant comparative analysis whereby groups are compared on the basis of theoretical similarities and differences.

In contrast to traditional deductive research that usually begins with an existing theory and then develops and tests hypotheses derived from that theory to determine support, inductive research begins with information-gathering, coding, and iterating between emerging theoretical ideas in the data and the literature to refine or build theory (Strauss & Corbin, 1990). This theoretical framework requires researchers to code as close to the data as possible. Recognizing that several ideas can be communicated in one spoken (and transcribed) sentence, researchers using this method first engage in open coding, or a coding ritual called "in vivo," based directly on words used by participants and are as similar to the terms used in the data as possible without applying external theoretical framings. Such line-by-line coding requires the researchers to examine individual words and phrases (Locke, 2001). Next, in a process analogous to creating superordinate categories in traditional content analyses, researchers combine the first-order codes into second-order codes and then generate aggregate dimensions informed by the literature to develop a model of the target topic. This approach is generally used in studies that do not attempt to quantify the textual data (e.g., create ratings, scores, or word frequencies), but instead try to summarize the content or generate theory. For example, using in-depth interviews with 29 US Navy couples, Beckman and Stanko (2019) were able to extend boundary theory by discovering the types of relational boundary work practices couples engaged in individually and collectively to build their resilience as a couple.

Of the 341 articles in this review, 171 (50.2%) used both qualitative and quantitative analyses, which involved converting text to quantitative data. A common example was counting the number of words in the text that represented a certain sentiment (e.g., positive or negative tone). These estimates then were used to correlate with or predict important outcomes. For example, Love, Lim, and Bednar

(2017) analyzed the sentiment of 200 news articles on CEOs and used this variable ("CEO media tenor") to predict firm reputation based on Fortune Magazine's "Most Admired Companies" surveys. They took a number of steps to create a quantitative measure of tenor. First, using LIWC's dictionaries of positive and negative words, they counted the number of positive and negative words in each response. Then, they created ratios of positive to negative words used and negative to positive words used. Finally, they coded articles as positive if the positive ratio was 0.65 or greater and negative if the negative ratio was 0.65 or greater. This is only one way of creating a sentiment score.

In a similar study on media favorability of organizations, Bednar (2012) also used LIWC to evaluate news articles, but calculated the mean frequency of positive and negative words "from all articles about a sample firm in a given year" (p. 138). Bednar tested whether media favorability was affected by formal board independence and also whether media favorability affected CEO pay and likelihood of CEO dismissal. Scholars have also measured sentiment as a simple percentage of the number of sentiment words (positive or negative) to the total number of words in a text. Wilson, DeRue, Matta, Howe, and Conlon (2016) applied this method in a study of emotional displays in negotiations. Instead of creating positive-to-negative ratios, and vice versa, Wilson et al. created an index of the percentage of positive emotions (e.g., words like "agree," "great," or "nice") per transcript. They then tested whether personality similarity between negotiators enhanced positive emotional displays and whether positive emotional displays quickened agreement time and reduced perceptions of relationship conflict. They found support for these hypothesized relationships.

Finally, of the 341 articles, 17 (5.0%) reviewed or promoted CATA. The reviews had somewhat different purposes, the most common of which were: (1) introducing text mining (e.g., Luciano, Mathieu, Park, & Tannenbaum, 2018), (2) providing a broad overview of CATA (e.g., Short et al., 2018), and (3) presenting specific techniques that included CATA (e.g., Crayne & Hunter, 2018; Janasik, Honkela, & Bruun, 2009; Shortt & Warren, 2019; Slutskaya, Game, & Simpson, 2018).

In sum, there were three types of CATA studies in the management literature. Nearly half used purely qualitative data to develop new theories. Nearly half included both qualitative and quantitative data, usually by using dictionaries to quantify the text data and predict outcomes. Finally, nearly 5% reviewed or promoted CATA.

Types of Textual Data

Perhaps what makes CATA so promising for researchers is that any and all written text can be analyzed. The text data used in the literature included various types collected intentionally for the study (e.g., interviews and open-ended responses on surveys) and types of existing data created for other purposes (e.g., letters to shareholders, news articles and press releases, and online reviews and social media). Of the 324 non-review articles, nearly half (45.4%) used secondary data only, about a third (36.7%) used primary data only, and 17.9% used both.

The most common source of text data were transcripts from interviews, focus groups, speeches, phone calls, and others with 165 studies (50.6%) in our review using these. Most of these data were collected by the researchers themselves.

Illustrative attributes from primary sources of text data included perceptions of fit (Chuang, Hsu, Wang, & Judge, 2015), humble leadership (Owens & Hekman, 2012), cognitions and emotions (Zuzul, 2018), and identity (Creed, DeJordy, & Lok, 2010; Gioia, Price, Hamilton, & Thomas, 2010). Other primary data sources included observations (e.g., researcher recordings of behavior based on visual observations) (51; 15.7%) and open-ended responses (18; 5.6%). Illustrative attributes from these types of text data include team cognitive maps (Carley, 1997), boundary management tactics (Kreiner, Hollensbe, & Sheep, 2009), and positive–negative sentiment (Liang et al., 2016).

In terms of existing data, the most commonly used text sources were types of organizational reports (e.g., letters to shareholders) (55; 17.0%) followed by archival data (48; 14.8%), and news-related documents (e.g., press releases) (47; 14.5%). Examples of attributes measured by these sources were CEO characteristics such as narcissism (Buyl, Boone, & Wade, 2019) and entrepreneurial orientation (Engelen, Neumann, & Schmidt, 2016), or firm-specific attributes such as organizational culture (Pandey & Pandey, 2017) and organizational values (Kabanoff & Holt, 1996; Kabanoff, Waldersee, & Cohen, 1995).

The next most common existing sources were journal articles and abstracts (15; 4.6%), and web-related content (15; 4.6%). For example, Antons, Joshi, and Salge (2018) text mined the topics and rhetorical features of more than 1,600 management journal articles to predict their scientific impact. Meanwhile, Schmiedel, Müller, and vom Brocke (2019) text mined online reviews of almost 300,000 Fortune 500 companies to examine organizational culture. Scholars have also examined the sentiment of online content (Barlow, Verhaal, & Hoskins, 2018). Forty articles (12.4%) used other types of existing textual data that did not fall into previous categories such as job postings, cases, mission statements, and crowdfunding campaigns.

As noted, almost half of the studies in our review used existing text data not originally collected for the purpose of text analysis research (e.g., organizational annual reports). One way to interpret this finding is that CATA is an unobtrusive and clever way to assess previously untapped areas due to difficulty in data collection. This may lead to the identification and measurement of new constructs or the refinement of existing constructs (Newman, Harrison, Carpenter, & Rariden, 2016). It may also help avoid the persistent faking and impression management common in data collected for personnel selection (e.g., personality indices). For example, some scholars have suggested that a candidate's social media (e.g., Facebook) could be particularly useful for recruiters because it may represent a more honest presentation of the candidate's true personality compared to interview responses (Hartwell & Campion, 2019).

However, secondary data sources come with notable limitations that create concerns about measurement validity because the data are not collected for the purposes of the research. To maximize utility of CATA-specific data collection, we should not solely rely on the text data we normally collect (e.g., asking for "any other comments"), but researchers should instead purposively collect text data to measure the intended or desired constructs. That is, if information is sought on a specific attribute, the questions should ask specifically about that attribute rather than inferring it from information collected for other purposes. For example, if the goal is to measure past leadership accomplishments, it is better to ask about past leadership accomplishments and text mine the responses than to text mine indirect text data such as personal statements. This is because the direct approach is more likely to solicit relevant and complete data on leadership, while the personal statement may focus on life goals and tangential constructs and not past accomplishments.

In summary, CATA can be used to analyze virtually any type of textual data and allows for researchers and organizations alike to exploit sometimes ignored or overlooked types of data (e.g., comments in worker engagement surveys). Researchers may be able to identify new constructs or refine existing ones through text data. Text data may also help researchers avoid some of the problems with purposefully collected data (like impression management). However, if text data are collected to measure a particular attribute, researchers should ask about that attribute directly as opposed to inferring it from other indirect information.

Types of Software

Types of software and the frequency with which they were used in studies in this review are listed in Table 1, from most common to least common. Only software used twice or more were listed for parsimony. The remaining software used only once in the review are grouped as "Only Used Once." It is important to note that some research teams used more than one type of software in their text analysis. The most commonly used types of software for CATA were those developed for researcher-intensive coding rather than automation. Nvivo (138; 42.6%) and Atlas.ti (30; 9.3%) were used in more than half of the studies. Researchers often use these programs to engage in inductive coding (see Type of Studies section above).

A similarly popular type of software was dictionary-based programs such as LIWC (82; 25.3%) and DICTION (27; 8.3%). In all, dictionary-based software was used in more than a third of the studies (38.0%). While LIWC was used most often to assess sentiment of text, it, along with DICTION and the others, have also been used to capture content such as entrepreneurial orientation (Short, Broberg, Cogliser, & Brigham, 2010). This type of software has been available publicly for several decades and has since undergone a number of iterations to accommodate changes in language or cultural modifications of word sentiment and meaning (Pennebaker, Chung, Ireland, Gonzales, & Booth, 2007). As evidenced by our review, LIWC has, by far, been the most popular of these types of software and the development of which we will discuss here to present a clearer understanding of its validation and how it works.

Beginning with general word generation, researchers amassed a large database of words and resources (e.g., English dictionaries, Roget's Thesaurus, affects scales). Next, they solicited feedback from SMEs (or "judges" as they are referred to in the manual) to provide their expertise regarding what should and should not be included in the dictionaries. Finally, they evaluated the psychometrics of the software, considering word use frequency and other elements. Information on its psychometric properties is readily available (Pennebaker et al., 2007). In addition to sentiment, LIWC has also been used to score personality and motivations. For example, Hirsh and Peterson (2009) found several linguistic correlates of the Big Five: conscientiousness was associated positively with achievement-related words; neuroticism was associated with anxiety and negative emotions; extraversion and agreeableness were associated with family related words or those that represented interpersonal concern; and openness was associated with words related to perceptual processing such as hearing and seeing.

Much like LIWC, DICTION is a particularly flexible program that allows for both content and sentiment analyses. However, the built-in dictionaries were developed specifically to extract only the following qualities: certainty, activity, optimism, realism, and commonality (DICTION, 2019). Yet, researchers are not limited to these attributes and can generate and use their own dictionary(ies) in DICTION. For example, in their study on leadership rhetoric, Bligh, Kohles, and Meindl (2004) used DICTION's optimism dictionary and developed five dictionaries in addition to optimism (collectives, faith, patriotism, aggression, and ambivalence) to determine attributes of leaders from presidential speeches. The remaining dictionary-based software used in the literature we reviewed included ones like Cat Scanner (5; 1.5%) and MonoConc Pro (5; 1.5%) that analyze text data similarly to LIWC and DICTION by counting the number of words per response that exist in a construct's dictionary.

The more sophisticated software used in our literature review included R (7; 2.2%), Python's various toolkits (4; 1.2%) and SPSS Modeler (3; 0.9%). With these software, researchers are able to apply advanced analytics such as NLP. As noted previously, NLP (and computational linguistics) considers relationships among words (using n-grams or strings of words) as well as the words themselves. It also includes preprocessing by eliminating stop words (e.g., the, a, in), *stemming* (reducing words to core terms like "working" and "worked" to "work"), and *lemmatizing* (converting inflected forms to a common term, such as converting "is" and "am" to "be"). NLP assumes that words close in meaning will occur close together in text.

There are a couple additional important points to make regarding software. First, 22 (6.8%) articles did not disclose their software and 25 (7.7%) used software that no other study used. Of these 25, several were not commonly recognizable partly because they were homemade by researchers or have become outdated (e.g., Recursive Inspection of Text Scan) with the advent of more advanced systems. Second, only 22 studies (6.8%) in our review used more sophisticated software suggesting that as a field we are underutilizing advanced techniques. Furthermore, few used commercial software products such as SPSS. This is surprising because commercial products are more complete, have better documentation, and are more user friendly. This may be because such products are relatively new on the market or because they are still fairly expensive (but trending downward). Regardless, the important observation is that the software used by management scholars to this point has been the less sophisticated software. The field has relied on rather simplistic applications of CATA based largely on word counts or significant human judgment (especially for content analysis). This may be limiting the sophistication of possible CATA research. Using more advanced approaches may allow the researcher to derive more construct validity information and broaden the range of applications.

Taken together, management researchers have at least eight choices of commonly used and trusted software to perform CATA, which can be divided into four types: (1) software to support manual content analysis (Nvivo and Atlas), (2) word dictionaries/sentiment analyses software (LIWC and DICTION), (3) programing software and languages (R and Python), and (4) commercial predictive analytics packages that include text mining modules (SPSS and SAS). Therefore, we researched and evaluated those eight software packages. We review several considerations likely to be important to researchers and summarize our findings on software in Table 2. The information presented was obtained in 2019 and may have been updated since in terms of both features and price.

Table 2. Content Analysis Software.

Software to Support Manual Content Analyses		
Software Name	**Nvivo** (earlier version called “Nud.ist”, or Non-numerical unstructured data indexing, searching, and theorizing)	**Atlas.ti**
Source to acquire & cost	Access website here: https://www.qsrinternational.com/nvivo/home Cost information: Academic license is $700 (Pro) & $800 (Plus) Government license is $979 (Pro) & $1119 (Plus) Commercial $1399 (Pro) & $1599 (Plus)	Access website here: https://atlasti.com/ Cost information: Academic license for single user is $750 and increases incrementally depending on number of user licenses (e.g., $3,000 for 5; $5,7630 for 10); Government license for single user is $1,290 and increases incrementally depending on number of user licenses (e.g., $4,650 for 5; $9,300 for 10); Commercial license for single user is $1,840 and increases incrementally depending on number of user licenses (e.g., $6,800 for 5; $13,200 for 10)
Function	Generally used for inductive content analyses; humans generate their own codes for the text data and assign those codes to words, phrases, or lines of data; software keeps track of codes; not capable of more advanced modeling.	Generally used for inductive content analyses; humans generate their own codes for the text data and assign those codes to words, phrases, or lines of data; software keeps track of codes; not capable of more advanced modeling.
Information on efficacy (e.g., validity)	While it has the capability of extracting its own categories, this function is largely based on frequency of words; otherwise, categories are usually human-developed, and psychometric and SME agreement analyses are conducted outside the software.	Categories are entirely human-developed, and psychometric and SME agreement analyses are conducted outside the software.

Table 2. *(Continued)*

	Software to Support Manual Content Analyses	
Information on usability (e.g., difficulty to learn; technical support)	Relatively intuitive to use (used in almost half of the studies in our review); technical support appears readily available and there are online trainings on how to use the software, as well as videos on YouTube.	Relatively intuitive to use and though only used in ~9% of studies in our review, it is a recognized software for inductive coding in management research; technical support appears readily available and there are free and for-cost training available, as well as videos on YouTube.
Likely applicability to Human Resources (HR) data	Can be used to analyze textual data of all types.	Can be used to analyze textual data of all types.
Likely applicability for measuring human attributes relevant to HR systems	Because categories are developed by humans, this software is flexible for researchers to code whatever they want to measure.	Because categories are developed by humans, this software is flexible for researchers to code whatever they want to measure.
Anticipated advantages	Easy to learn, low cost, widely used, and flexible to meet researcher coding needs.	Easy to learn, low cost, widely used, and flexible to meet researcher coding needs.
Anticipated disadvantages or problems	Validity and reliability studies occur outside of software; not capable of more advanced modeling.	Validity and reliability studies occur outside of software; not capable of more advanced modeling.
Overall evaluation for potential usefulness	Highly useful software if the researcher's goal is to conduct manual content analysis. However, it does not automate the content analysis process in the sense of extracting and scoring the content.	Highly useful software if the researcher's goal is to conduct manual content analysis. However, it does not automate the content analysis process in the sense of extracting and scoring the content.
	Word Dictionaries / Sentiment Analyses Software	
Software Name	**LIWC** (Linguistic Inquiry Word Count)	**DICTION**
Source to acquire & cost	Access website here: http://liwc.wpengine.com/ Cost information: Academic license is $89.95; commercial version available, but must apply through Receptiviti. The manual also states that any commercial use of the LIWC dictionaries is forbidden without permission through Receptiviti.	Access website here: https://www.dictionsoftware.com/ Cost information: Academic license is $219 and corporate license is $269.
Function	Originally used to analyze sentiment (affect) from text using in-house, pre-validated dictionaries. It has about 50 dictionaries in 7 topic categories of psychological processes as well as a number of other language metrics (e.g., number of types of words used). They are described in the manual. It can also be used to analyze content of text. Allows for the use of a researcher's own custom dictionary, but not capable of more advanced modeling.	Can be used for both content and sentiment; has five in-house dictionaries (certainty, activity, optimism, realism, and commonality); allows for the use of a researcher's own custom dictionary, but not capable of more advanced modeling.

Table 2. *(Continued)*

Word Dictionaries / Sentiment Analyses Software		
Information on efficacy (e.g., validity)	Fairly transparent in the development and refinement of their dictionary; internal consistencies are in the Language Manual.	No publically available information about the development and refinement of the dictionaries.
Information on usability (e.g., difficulty to learn; technical support)	It is very easy to learn and use, which probably explains its prevalence in our review (used in ∼25% of studies). Technical manual (Operator Manual) is readily available and technical support is provided by the researchers who developed the software. Also some videos on YouTube.	Also easy to use. Not as well-known as LIWC, but still recognized (∼8% of studies in the review). Technical manuals are available (Manual DICTION and Using DICTION) and technical support is available via email or phone as well as videos on YouTube.
Likely applicability to Human Resources (HR) data	Can be used to analyze textual data of all types.	Can be used to analyze textual data of all types.
Likely applicability for measuring human attributes relevant to HR systems	The 50 dictionaries measure human attributes, many of which could be relevant to HR management purposes. It is also possible for researchers to develop their own dictionaries.	The 5 existing dictionaries measure human attributes, some of which could be relevant to HR management purposes. It is also possible for researchers to develop their own dictionaries.
Anticipated advantages	Straightforward, validated, and recognized software. Has a large number of in-house dictionaries. Easy to learn and use.	Seemingly straightforward and recognized software. Easy to learn and use.
Anticipated disadvantages or problems	Best at measuring sentiment only and not capable of doing more advanced machine learning procedures.	Not transparent about validity and reliability of dictionaries. Only five in-house dictionaries. Best at measuring sentiment only and not capable of doing more advanced machine learning procedures.
Overall evaluation for potential usefulness	May be useful if the purpose is to measure sentiments. Because it cannot do more advanced machine learning, its role will only be adding the sentiment analysis. This is the most fully developed and comprehensive set of measures of various sentiments.	May be useful if the sentiments that it contains are relevant to the research. Because it cannot do more advanced machine learning, its role will only be adding the sentiment analysis.
Programing Software/Language		
Software Name	**R** (The R Project)	**Python**
Source to acquire & cost	Access website here: https://www.r-project.org/ It is free	Access website here: https://www.python.org/ It is free
Function	R is a highly powerful and flexible software. It is popular among data scientists and more recently (past 10 years or so) has been used by management scholars. It is open-sourced, meaning code is made freely available by other researchers. It can conduct traditional statistical analyses, content and sentiment analyses, text mine, and is capable of advanced modeling.	It is a programing language. It is highly powerful and flexible and can be used for myriad tasks including data analyses, text analysis and machine learning, and developing websites/ mobile apps. Need to download packages to be able to run analyses. It is open-sourced and packages are available on Python's website, as well as other places online.

Table 2. *(Continued)*

	Programing Software/Language	
Information on efficacy (e.g., validity)	Can incorporate pre-validated dictionaries to conduct analyses and validity and reliability analyses can be done using R.	Can incorporate pre-validated dictionaries to conduct analyses and validity and reliability analyses can be done using Python.
Information on usability (e.g., difficulty to learn; technical support)	As it requires basic understanding of software coding, it may be difficult to learn. Because it is a free, open-sourced software, no person or team is available to answer questions. The first line on the software's help page is "Before asking others for help, it's generally a good idea for you to try to help yourself," and they provide a few ways in which researchers can do that. Otherwise, materials abound from books (recommended: Discovering Statistics Using R by Field, Miles, & Field, 2012; Silge & Robinson, 2017) to websites (e.g., Stack Overflow; GitHub, coding from NLP seminar at a professional conference available here: https://github.com/coryamanda/SIOP2019_NLP_organization_Research), as well as videos on YouTube.	As it is a programing language and requires basic understanding of software coding, it may be difficult to learn. Tutorials are available. There are many guides online, as well as books, and there is an email for concerns not addressed on the site. Many packages can be found on GitHub and videos can be found on YouTube.
Likely applicability to Human Resources (HR) data	Can be used to analyze textual data of all types.	Can be used to analyze textual data of all types.
Likely applicability for measuring human attributes relevant to HR systems	Because researchers are able to develop their own measures, it is possible to use R to measure human attributes relevant to HR management.	Because researchers are able to develop their own measures, it is possible to use Python to measure human attributes relevant to HR management.
Anticipated advantages	Flexible and powerful and some coding available from others.	Flexible and powerful and some coding available from others.
Anticipated disadvantages or problems	Unless you understand the minutiae of software coding, you run the risk of conducting the wrong analyses, particularly if you simply use or repurpose someone else's coding. It will likely be time-consuming to learn, especially for nonprogrammers.	Unless you understand the minutiae of software coding, you run the risk of conducting the wrong analyses, particularly if you simply use or repurpose someone else's coding. It will likely be time-consuming to learn, especially for nonprogrammers.
Overall evaluation for potential usefulness	Could be very useful, but complexity and time to learn may be too extensive for the casual user.	Could be very useful, but complexity and time to learn may be too extensive for the casual user.

Table 2. *(Continued)*

Commercial Predictive Analytics Software Packages		
Software Name	SPSS Modeler Premium	SAS Enterprise Miner
Source to acquire & cost	Access website here: https://www.ibm.com/products/spss-modeler Cost information: Free 30-day trial, $199/month subscription, $7,430 for Professional license, $12,400 for Premium license, and $25,600 for Gold, but Premium or Gold needed for text mining capability. Gold includes "Collaboration & Deployment Services for model deployment and management."	Access Text Miner website here: https://www.sas.com/en_us/software/text-miner.html Access Enterprise Miner website here: https://www.sas.com/en_us/software/enterprise-miner.html Text Miner is a component of Enterprise Miner, which allows for machine learning. Seems to require two licenses with Enterprise Miner downloaded first. Pricing appears to be based on the intended user and uses, but likely to be the most expensive option.
Function	It is a user-friendly, powerful, and flexible software. It allows for more advanced features such as concept extraction and text mining, categorizations, and machine learning. It allows extensions to embed R and Python code into an SPSS modeler stream by simply inserting a "node" (analogous to a plugin). This allows running R and python scripts to import data, apply transformations, build and score models, display outputs, and export data. Plus, as part of a predictive analytics suite, it includes a very wide range of statistical procedures that are preprogrammed and easy to apply.	It is user-friendly, powerful, and flexible software. It allows for more advanced features such as concept extraction and text mining, categorizations, and machine learning. Plus, as part of a predictive analytics suite, it includes a very wide range of statistical procedures that are preprogrammed and easy to apply.
Information on efficacy (e.g., validity)	Has several built-in dictionaries, but there is little if any detail on the psychometric properties of the dictionaries because it is deemed "proprietary." Researchers are able to develop and validate their own dictionaries or add concepts to existing ones. Includes sentiment analyses, but in-house sentiment dictionaries are very basic and limited.	It appears SAS does not use a dictionary-based approach, but instead "relies primarily upon pattern recognition" (p. 10; see http://opim.wharton.upenn.edu/~sok/papers/s/sas/wp_3633.pdf). It also appears that SAS Sentiment Analysis is a separate software.
Information on usability (e.g., difficulty to learn; technical support)	It is user friendly because it is a point-and-click software. Because most researchers in management use SPSS, the interface will not require much acclimation. There are also many avenues of support including IBM technical support, videos on YouTube, and many other online resources.	It is user friendly because it is a point-and-click software. Because many researchers in management use SAS, the interface will not require much acclimation. There are many resources through SAS, as well as videos on YouTube and other online guides.

Table 2. *(Continued)*

	Commercial Predictive Analytics Software Packages	
Likely applicability to Human Resources (HR) data	Can be used to analyze textual data of all types. Also able to incorporate quantitative variables simultaneously using a wide range of built in predictive models.	Can be used to analyze textual data of all types. Also able to incorporate quantitative variables simultaneously using a wide range of built in predictive models.
Likely applicability for measuring human attributes relevant to HR systems	Because researchers are able to develop their own measures, it is possible to measure human attributes relevant to HR management.	Because researchers are able to develop their own measures, it is possible to measure human attributes relevant to HR management.
Anticipated advantages	Flexible and powerful, straightforward, familiar, and support is available.	Flexible and powerful, straightforward, familiarity, and support is available.
Anticipated disadvantages or problems	Cost is greater than other software.	Cost is greater than other software.
Overall evaluation for potential usefulness	Probably the easiest way for management researchers who are not programmers to be able to conduct the more complex types of CATA (those involving NLP and LSA), but the cost of the software may be significant for the casual user.	Probably the easiest way for management researchers who are not programmers to be able to conduct the more complex types of CATA (those involving NLP and LSA), but the cost of the software may be significant for the casual user.

RESEARCH RELEVANT TO EMPLOYMENT

Of particular relevance to the employment literature are those studies where researchers measured human attributes applicable to staffing decisions (e.g., various types of knowledge, skills, abilities, and other characteristics, or knowledge, skills, abilities, and other characteristics or [KSAOs]). We reexamined the 324 empirical articles in our review and found 66 articles that used CATA that bear on employment practices, which we classified into recruitment, selection, performance management, engagement, leadership, and turnover. We review how these articles can contribute to each HR practice, especially focusing on the implications of using CATA for measuring attributes relevant to employment decisions. Details of these studies (i.e., attributes measured, type of text data, and software used) are presented in Table 3. Finally, we elucidate a number of important takeaways about using CATA for employment through an analysis of these 66 studies.

RECRUITMENT

Two studies used CATA to look at recruitment. In the first, Peltokorpi and Vaara (2014) conducted a qualitative study on language-sensitive recruitment and knowledge transfer in multinational corporations (MNC) using Nvivo. Language-sensitive recruitment refers to “recruitment practices in which a certain proficiency in the corporate language is used as a precondition for employment”

Table 3. Employment-related Articles Using CATA.

HR Practice	Authors (Year)	Attribute(s) Measured	Type of Text Data	CATA Software
Recruitment	Banks et al. (2019)	Recruitment signals	Websites	DICTION
	Peltokorpi and Vaara (2014)	Language-sensitive recruitment and knowledge transfer	Interviews	Nvivo
Selection	Anglin, Wolfe, Short, McKenny, and Pidduck (2018)	Psychological capital	Online crowdfunding campaigns	DICTION
	Anglin et al. (2018)	Narcissistic rhetoric	Online crowdfunding campaigns	DICTION
	Basu and Savani (2017)	Depth of cognitive processing	Open-ended survey questions	LIWC
	Campion et al. (2016)	Job-related skills: communication, critical thinking, people, leadership, managerial, and factual knowledge	Assessment records	SPSS Modeler
	Chen and Latham (2014)	Achievement orientation	Open-ended survey questions	LIWC
	Clark et al. (2018)	Cognitive effort	Think-aloud protocol	Nvivo
	Connelly, Zweig, Webster, and Trougakos (2012)	Knowledge hiding	Interviews	Nvivo
	Corner and Ho (2010)	Opportunity identification	Interviews	Nvivo
	Crilly (2017)	Analytic thinking and cognitive processing	Open-ended survey questions	LIWC
	Gibson et al. (2011)	Virtual job characteristics	Interviews	Atlas.ti
	Josefy et al. (2017)	Entrepreneurial orientation	Online crowdfunding campaigns	Unclear, but used Short et al. (2010) entrepreneurial orientation dictionary
	Kellogg, Orlikowski, and Yates (2006)	Boundary-spanning practices	Interviews and observations	Nvivo
	Kobayashi, Mol, Berkers, Kismihok, and Den Hartog (2018a)	Job task information	Job descriptions	R
	Kobayashi, Mol, Berkers, Kismihok, and Den Hartog (2018b)	Worker attributes of job descriptions	Job descriptions	R
	Laureiro-Martinex and Brusoni (2017)	Cognitive flexibility	Open-ended responses to cognitive tasks	Nvivo
	Madera et al. (2009)	Communal and agentic language	Letters of recommendation	LIWC
	Martin (2016)	Organizational values	Handwritten stories	LIWC

Table 3. *(Continued)*

HR Practice	Authors (Year)	Attribute(s) Measured	Type of Text Data	CATA Software
	Moore, Lee, Kim, and Cable (2017)	Self-verification striving	Interviews	LIWC
	Moss et al. (2015)	Virtuous orientation and entrepreneurial orientation	Load descriptions	LIWC
	Moss, Renko, Block, and Meyskens (2018)	Social value orientation and economic value orientation	Microenterprise narratives	LIWC
	Nadkarni and Narayanan (2005)	Cognitive structure of cognitive ability	Case analyses	Netanalysis
	Reinecke and Ansari (2015)	Clock-time orientation	Interviews, observations, and archival documents	Nvivo
	Shantz and Latham (2009)	Achievement orientation	Handwritten stories	LIWC
	Watson, Dada, Wright, and Perrigot (2019)	Entrepreneurial orientation rhetoric	Organizational narratives	DICTION
	Waung et al. (2017)	Impression management tactics	Resumes and cover letters	LIWC
	Wilhelmy, Kleinmann, König, Melchers, and Truxillo (2016)	Interviewer impression management	Interviews, observations, memos, and archival documents	Atlas.ti
Performance Appraisal	Carley (1997)	Team cognitive maps	Open-ended survey responses	Automap
	Smither and Walker (2004)	Favorability/unaffordability; Trait/behavior focus	Comments from 360-degree surveys	Did not report
	Speer (2018)	Sentiment	Performance management narrative comments	R
Engagement	Barley et al. (2011)	Emails and employee overload	Interviews	Atlas.ti
	Petriglieri (2015)	Identity reconstruction	Interviews	Atlas.ti
	Reay et al. (2017)	Changes in professional role identity	Interviews	Nvivo
	Stanko and Beckman (2015)	Boundary control efforts	Interviews, archival data	Nvivo
Leadership	Akinola et al. (2018)	Positive/negative affect of delegating tasks	Open-ended survey questions	LIWC
	Bligh et al. (2004)	Rhetorical characteristics (optimism, collectives, faith, patriotism, aggression, & ambivalence)	Presidential speeches	DICTION
	Boling et al. (2015)	Entrepreneurial orientation	Letters to shareholders	LIWC
	Buyl et al. (2019)	CEO entrepreneurial orientation	Letters to shareholders	LIWC
	Chatman, Caldwell, O'Reilly, and Doerr (2014)	Firm adaptability	Other ratings of annual reports	LIWC
	Detert and Treviño (2010)	Leader influence on employee voice	Interviews	Nvivo
	Engelen et al. (2015)	Entrepreneurial orientation	Letters to shareholders	DICTION
	Engelen et al. (2016)	Entrepreneurial orientation	Letters to shareholders	DICTION

	Fanelli, Misangyi, & Tosi (2009)	CEO charismatic visions	Letters to shareholders	DICTION
	Grühn, Strese, Flatten, Jaeger, and Brettel (2017)	CEO entrepreneurial orientation change	Letters to shareholders	LIWC
	Harrison et al. (2019a)	Big Five Personality	CEO earnings conversation transcripts	R
	Harrison et al. (2019b)	Big Five Personality	CEO earnings conversation transcripts	R
	Kang and Kim (2017)	CEO narcissism	News articles	LIWC
	Keil, Maula, and Syrigos (2017)	CEO entrepreneurial orientation	Letters to shareholders	Unclear, but adopted Short et al. (2010) entrepreneurial orientation dictionary
	Lanaj et al. (2019)	Clout	Open-ended survey questions	LIWC
	Liang et al. (2016)	Sentiment	Open-ended survey questions	LIWC
	Malhotra et al. (2018)	CEO extraversion	CEO responses in quarterly earnings conference calls	LIWC
	McAlearney (2006)	Leadership development challenges	Interviews	Atlas.ti
	McKenny et al. (2013)	CEO psychological capital	Letters to shareholders	DICTION
	Osborne, Stubbart, and Ramaprasad (2001)	CEO mental models	Letters to shareholders	WordCruncher
	Owens and Hekman (2012)	Humble leadership	Interviews	Atlas.ti
	Ridge and Ingram (2017)	Top management team modesty	Conference call transcripts	Nvivo; DICTION
	Short et al. (2010)	Entrepreneurial orientation	Annual reports	DICTION
	Sonenshein (2014)	Creative use of resources	Interviews and archival documents	Nvivo
	Titus, Parker, and Covin (2019)	CEO entrepreneurial orientation	Letters to shareholders	LIWC
	Toegel, Kilduff, and Anand (2013)	Emotion helping	Interviews	Nvivo
	Wolfe and Shepherd (2015)	Entrepreneurial orientation	Letters to shareholders	LIWC
Turnover	Felps et al. (2009)	Job search behaviors	Focus groups	Atlas.ti
	Follmer et al. (2018)	Managing perceptions of misfit	Interviews	Nvivo
	Rothausen et al. (2017)	Identity predictors of voluntary turnover	Interviews, focus groups	Nvivo
	Sajjadiani et al. (2019)	Attributions for turnover	Open-ended survey responses	SPSS Modeler

(Peltokorpi & Vaara, 2014, p. 601). Analyzing interview transcripts of more than 130 MNC employees, they found that a focus on language proficiency in hiring can aid in knowledge transfer as well as make communication and network building easier. However, they also found that it can be counterproductive to knowledge transfer due to weaker host-country embeddedness and identification. In the second recruitment paper, Banks et al. (2019) tested how the strength of recruitment signals varied across domestic and international organizational websites. They text mined and analyzed data from 162 organizational websites across 21 countries using NLP to identify themes. They found that MNCs tend to standardize across countries and that differences did not depend on culture.

These two studies showed how text mining can help to improve recruiting messages, including features of recruiting messages that may not have been observed without text analysis. Using CATA to understand recruitment signals may also make culture differences in recruitment methods easier to assess.

SELECTION

Of the 66 studies, 28 were in the context of, or directly relevant to, selection. These could be additionally broken into three subcategories of attributes measured: individual skills, personality or orientations, and organizational or job characteristics. Skills were evaluated in a number of ways. For example, Campion, Campion, Campion, and Reider (2016) used text mining with NLP to measure several job-related skills (communication, critical thinking, people, leadership, managerial, and factual knowledge) based on accomplishment records in a hiring context. Clark, Li, and Shepherd (2018) measured cognitive effort with think-aloud protocols, and Crilly (2017) used open-ended survey responses to capture analytic thinking and cognitive processing. Finally, Nadkarni and Narayanan (2005) examined cognitive structure and cognitive ability by using network analyses.

Several studies also examined orientations or personality. For example, Josefy, Dean, Albert, and Fitza (2017), assessed entrepreneurial orientation of aspiring entrepreneurs in online crowdfunding campaigns. These scholars used the entrepreneurial orientation dictionary originally developed by Short et al. (2010) (see Leadership section below). Moss, Neubaum, and Meyskens (2015) also used this dictionary to assess entrepreneurial orientation, but did so with loan descriptions. Similarly, Chen and Latham (2014) and Shantz and Latham (2009) measured achievement by using the achievement dictionary in LIWC to analyze open-ended survey questions and handwritten stories, respectively. Others in this category included Madera, Hebl, and Martin (2009) who measured characteristics of candidates (communal versus agentic traits) in letters of recommendation using LIWC, and Waung, McAuslan, DiMambro, and Miegoć (2017) who assessed the types of impression management tactics candidates used in resumes and cover letters. Waung et al. found evidence of eight impression management categories: superlative use, adjective use, reference to fit, enhancement or entitlement, credit to external sources, individual ingratiation, institutional ingratiation, and outlook or values.

A few studies examined job or organizational characteristics relevant to employment. For example, Gibson, Gibbs, Stanko, Tesluk, and Cohen (2011) used the CATA to search for keywords in interview transcripts to capture variables related to virtual job characteristics (e.g., perceptions of electronic dependency and copresence, defined as "the subjective perception of closeness versus distance," p. 1484). In doing so, they expanded the traditional job characteristics model (Hackman & Oldham, 1975) to include virtual characteristics of modern work. Meanwhile, Martin (2016) examined how recent hires learned organizational values by using a "values" dictionary in LIWC to content analyze stories written by recent hires to illustrate the organization's values.

Taken together, results from these studies suggest that text data analyzed using CATA might be valuable for personnel selection. For example, this research identifies new constructs that may have implications for hiring such as communal traits that might predict team cohesion and agentic traits that might predict team performance (Madera et al., 2009). These studies also highlight how less traditional measures (e.g., storytelling) offer an opportunity to capture candidate personality traits without being as susceptible to social desirability in responses. Shantz and Latham (2009) also showed that we may be able to analyze stories and draw out achievement orientations regardless of writing quality. Finally, soliciting descriptions of past behavior from potential candidates that illustrates their values (or write descriptions of how they embody the company's values) could be a useful assessment tool (Martin, 2016).

PERFORMANCE MANAGEMENT

Three studies used CATAbin the context of, or directly relevant to performance management. In the first article, Smither and Walker (2004) analyzed the favorableness and the trait versus behavioral nature of comments in 360 ° surveys. They found that those who received a small number of unfavorable comments were able to improve their performance appraisal scores a year later. In the second study, Speer (2018) text mined performance management narrative comments using R to analyze the sentiment of the comments. Speer (2018) wrote, "The derived narrative scores were reliable across years, converged with traditional numerical ratings and explained incremental variance in future performance outcomes (performance ratings, involuntary turnover, promotions, and pay increases)" (p. 299). In the last article, Carley (1997) assessed team cognitive maps using open-ended survey responses and found that members of successful teams tend to have more commonly shared mental maps.

Taken together, two studies found great value in using the narrative comments associated with numerical performance management feedback that often go unanalyzed and may be more candid than numerical ratings. Speer's (2018) findings are particularly relevant to practice because examination of these non-quantitative data may help to improve criterion measurement during assessment development and validation. Moreover, Carley's study presents an alternative way to assess teams by identifying potential areas of conflict due to poor communication (poor shared mental model) enabling leaders to help members manage their team's performance. Finally, although Smither and Walker's study focused on how managers responded to narrative feedback, it suggests indirectly that managers may differ in their receptiveness to feedback, which could potentially be an attribute assessed at time of hire. It also demonstrates a possible use of an important performance appraisal technique that tends to solicit a significant amount of written text that can be analyzed and used more effectively and quickly with the help of CATA.

ENGAGEMENT

Four studies examined employee engagement using CATA. Two of the studies focused on organizational norms and policies. Barley, Meyerson, and Grodal (2011) assessed the impact of work emails on employee overload and found that there were a number of distinct norms and emotions interviewees felt about email such as feeling the need to respond immediately upon receiving an email, responding at night, and fear of falling behind, to name a few. Meanwhile, Stanko and Beckman (2015) analyzed how organizations exert control on employee technology use for non-work purposes to enhance workplace engagement. These scholars found that organizations enact policies both formal and informal to manage the boundaries between employees' work and home life in a sample of Navy personnel. In the other two studies relevant to engagement, the researchers examined identity. Petriglieri (2015) assessed how executives at British Petroleum reconstructed their organizational identities after the Gulf spill and found that they developed pathways to resolve ambivalence with their organization to successfully rebuild their organizational identities and re-engage. Meanwhile, Reay, Goodrick, Waldorff, and Casebeer (2017) studied how physicians changed and developed new organizational identities facilitated by other social actors within the organizations.

Because CATA can approach topics inductively, it allows scholars to generate frameworks regarding workplace cultures (norms) that could otherwise be neglected by imposing existing measures, thus pinpointing an important contribution of CATA to engagement research. Stanko and Beckman's study, in particular, demonstrated how using CATA can improve how we conceptualize certain HR practices. Their grounded theory approach allowed for the emergence of boundary control techniques centered on cultivating and maintaining employee attention that we would not have known otherwise had they used more top-down methods (e.g., established Likert-type measures). Further, measuring the existence of a policy and whether behaviors differ before and after implementation are an important part of the story. Stanko and Beckman were able to demonstrate *how* these policies changed employee behaviors providing additional insight into an important HR practice area.

LEADERSHIP

Of the 66 studies relevant to employment, 28 measured leadership variables or processes. Similar to CATA research on selection, the leadership research can be additionally broken down into subcategories: leader language, leader behaviors, and leader characteristics. Leader language included one study. In this study, Bligh et al. (2004) examined the rhetorical characteristics (optimism, collectives, faith, patriotism, aggression, and ambivalence) of 74 speeches by President George W. Bush before and after 9/11 and found that the rhetoric changed post-9/11.

Research on leader behaviors ranged from delegation and gender differences in sentiment regarding delegation (Akinola, Martin, & Phillips, 2018) to leader clout or confidence (Lanaj, Foulk, & Erez, 2019). For example, Liang et al. (2016) analyzed the sentiment of written descriptions of recalled interactions between subordinated and supervisors to assess abusive supervision. Further, Owens and Hekman (2012) conducted a grounded theory study to extract behaviors that represented humble leaders and found that "leader humility involves leaders modeling to followers how to grow and produce positive organizational outcomes by leading followers to believe that their own developmental journeys and feelings of uncertainty are legitimate" (p. 787).

Researchers have also used CATA to measure leader characteristics. The most popular characteristic, thus far, has been entrepreneurial orientation. Short et al. (2010) first used CATA to measure entrepreneurial orientation by developing a dictionary for the construct using CEO letters to shareholders. Many other research teams have utilized Short et al. (2010) entrepreneurial orientation dictionary or a modified version of it to assess this characteristic in CEOs or entrepreneurs (e.g., Boling, Pieper, & Covin, 2015; Engelen, Neumann, & Schwens, 2015; Engelen et al., 2016; Wolfe & Shepherd, 2015). Similarly, Harrison, Thurgood, Boivie, and Pfarrer (2019a; 2019b) illustrated how to use archival information to measure historically difficult-to-access samples (e.g., CEOs). In their studies, they developed an algorithm to assess the Big Five personality traits of CEOs and found that CEO personality was related to firm performance. Furthermore, Malhotra, Reus, Zhu, and Roelofsen (2018) measured the extraversion of CEOs based on their spoken responses to questions using LIWC and found that extraverted CEOs were more likely to acquire other companies.

Finally, researchers have also used CATA to understand leader influence. In their study, Detert and Treviño (2010) analyzed interviews with employees across organizational levels to understand how leaders (direct supervisors and skip-level supervisors) influence whether and how employees engage in voice behaviors. They found that direct supervisors influence employees' likelihood of speaking up at work, but also that skip-level managers additionally influenced the likelihood of speaking up, especially if an employee perceived that voicing an opinion would be futile. Often the distance between lower-level workers and upper-level managers appears too great of a distance to traverse for both parties. With differing goals and organizational knowledge, lower-level workers may feel unheard and upper-level managers may feel out-of-touch with what employees need or want.

Notably, not all leadership studies focused on leader behaviors or influence. One study used CATA to capture challenges to leadership development, which is relevant to organizations that run leader development programs or struggle with succession management. In this study, McAlearney (2006) used grounded theory to identify high-level leadership development challenges: industry lag, representativeness, professional conflict, time constraints, technical hurdles, and financial constraints.

In all, findings from research using CATA to measure leadership highlight potential uses for employment staffing as well as a way to enhance our basic understanding of leadership. For example, in their study, Malhotra et al. (2018) showed that extraverts used more words in spoken responses suggesting that answer length is a potential additional measure of extraversion in candidate assessments. Moreover, leaders appear to use certain types of linguistic features (e.g., Bligh et al., 2004; Lanaj et al., 2019) and it could be useful in research and practice to use automated text analysis to assess the rhetorical skills of applicants. Finally, several studies demonstrated alternative ways to assess personality that are less vulnerable to faking (e.g., Harrison et al., 2019a, 2019b).

TURNOVER

Four studies in our review examined constructs relevant to turnover. Three studies used a grounded theory approach while one used machine learning to develop a model to predict turnover. Follmer, Talbot, Kristof-Brown, Astrove, and Billsberry (2018) unfolded a process model of how perceptions of misfit influence adaptive strategies (e.g., voluntary turnover), while Rothausen, Henderson, Arnold, and Malshe (2017) examined how identity predicts turnover. Felps et al. (2009) analyzed focus group transcripts and generated a list of job search behaviors. They found that the frequency of words associated with job search behaviors was related to job embeddedness, commitment, and satisfaction, which suggests that word choices of employees (that could be collected in many ways) may predict individual turnover intentions. Finally, Sajjadiani, Sojourner, Kammeyer-Mueller, and Mykerezi (2019) trained a model to assess applicant work history from more than 35,000 resumes and open-ended responses using supervised machine learning to predict turnover among school teachers. This is a modern-day version of using weighted application blanks and biographical data, which has a history of validity for predicting turnover (e.g., Reilly & Chao, 1982).

Taken together, these studies demonstrate a number of important uses of CATA to help companies predict and manage turnover. For example, Rothausen et al. (2017) focus on procedures for validating qualitative research from Silverman and Marvasti (2008, pp. 257–270) (i.e., refutability, constant comparison, comprehensive data treatment, deviant-case analysis, and respondent validation) show that alternative types of validation appropriate for qualitative data may be very important if using CATA for hiring, given the need to demonstrate job relatedness. Furthermore, Felps's study shows how job search behaviors may augment the prediction of collective turnover beyond engagement survey ratings by analyzing the written comments. Finally, Sajjadiani et al. (2019) study shows how researchers and organizations have underutilized applicant work history as relevant to predicting turnover. It is likely that work history and other relevant applicant information are able to predict potential turnover at time of hire.

SUMMARY

We derived eight key takeaways on the use and value of CATA across HR practices.

Takeaway #1

CATA has helped identify new constructs that might be useful for building assessments for hiring decisions. This is not only because CATA allows for the analysis of new information (qualitative data), but also because CATA is typically used inductively with the expressed purpose of identifying unknown underlying constructs. For example, Short et al.'s (2010) development of an entrepreneurial orientation dictionary demonstrated that there were key linguistic features that differentiate those who are more or less entrepreneurially oriented. For additional examples, see Rothausen et al.'s (2017) work on identity and Follmer et al.'s (2018) work on perceptions of fit above.

Takeaway #2

CATA has made a significant impression in the leadership literature where 28 of the 66 studies in our review relevant to employment examined leadership. Each of these studies supports the notion that the language leaders use is an important influence mechanism. Similar to entrepreneurial orientation, the realization that there were linguistic differences among leaders and between leaders and subordinates is a notable contribution of CATA. Moreover, research on leadership and leader influence using grounded theory (e.g., Detert & Treviño, 2010) show that primary interviews and other qualitative data collection methods and using CATA to analyze the data can reveal important and previously ignored leader influences on employee behavior. Finally, studies such as McAlearney's (2006), which focused on challenges to leadership development, bear on criterion measurement for those in leadership positions. This suggests job performance for the managers in this organization (health care) can be constrained by a range of situational factors (Peters & O'Connor, 1980).

Takeaway #3

There has been much less focus on KSAs than Os. With few exceptions, CATA has been used to identify personality traits and behavioral orientations, with little research on using CATA to identify knowledge, skills, and abilities (for exceptions, see Campion et al., 2016; Nadkarni & Narayanan, 2005). This is a missed opportunity because KSAs are much more predictive of job performance (e.g., Schmidt & Hunter, 1998) and could potentially be measured well because KSAs are more definable, objective, and verifiable. There are also several other employment-related studies that allude to important employment-related considerations. Results from Gibson's study on virtual job characteristics could improve selection determinations as elements of virtual labor may be prohibitive for certain types of candidates. For example, job applicants may differ on their need for copresence, which has implications for their ability to manage the isolation often associated with virtual work.

The extraction of constructs using CATA also highlights the hierarchical nature of many construct domains. Nearly half of the CATA studies found that the constructs were hierarchically organized with primary, secondary, and even higher-levels of aggregation coding needed. It is well known that KSAOs are hierarchically organized (e.g., as recognized in the comprehensive taxonomies in O*NET; Peterson et al., 2001), and CATA can help researchers understand the hierarchical nature of new construct domains. Moreover, the hierarchical nature of KSAOs is important practically and theoretically because it means KSAOs are often highly related empirically and so precise matching to job requirements is less necessary than validation guidelines imply.

Takeaway #4

CATA exploits a range of data accessible and applicable to the employment context and can be done at a low cost. Specifically, CATA identifies a "new" type of data to be analyzed. Of course, text data are not "new," but are often considerably time-consuming to analyze and challenging to validate. As such, more automated text analysis methods allow researchers to take advantage of previously underutilized or unscored data such as narrative information in applications (e.g., personal statements, performance narratives). Because CATA provides an opportunity for organizations to exploit text data, organizations are able to derive more value from assessment information at a lesser cost. For example, Speer (2018), who text mined and analyzed the sentiment of performance management narrative comments astutely wrote:

> […] inclusion of narrative comments will result in (a) increases in total information and reliability, which is expected to occur across appraisal settings. In addition, in contexts where narratives are not explicitly linked to distributive outcomes (e.g., pay), narratives are likely to (b) exhibit a reduced amount of variance attributable to rater bias … it is pertinent to note that motivation to distort will also likely exist. For example, just as a lower traditional rating could lead to an unpleasant confrontation with a subordinate, negative comments could also spur unpleasant interactions that promote avoidance motivations … However, whereas this may occur for both mediums in likely equal probability, traditional narrative ratings are substantially more likely to be explicitly tied to distributive outcomes than narrative comments are, and therefore more likely to be affected by leniency bias. (pp. 304–306)

This chapter is particularly useful for employment because it demonstrates the potential to gather candid information in addition to ratings to improve job performance criterion measurement for assessment development and validation.

Furthermore, Campion et al. (2016) showed that training a model to score assessment records (e.g., narratives of past accomplishments) saved one organization at least $163,000 annually because it replaced human assessors. In this organization, three human assessors were used to score each candidate's narrative application information creating substantial costs during the selection process. By training a model against human ratings, Campion et al. (2016) illustrated how cost-saving CATA could be. Due to the reduced cost of evaluating narrative information, CATA may be a key to reducing adverse impact. Researchers can use CATA assessments that are valid yet show lower subgroup differences than mental ability tests, but are usually not used for large scale screenings of candidates due to high administrative costs, such as automated interviews and accomplishment records (Campion et al., 2016; Ployhart & Holtz, 2008) (see below for more information on CATA and adverse impact).

Takeaway #5

Scholars who have employed CATA have largely done so using very simple software, often, if not usually, conducting sentiment analysis based on existing dictionaries or manual content analysis with the assistance of software that merely manages and facilitates the manual sorting and coding of qualitative data (e.g., Nvivo or DICTION). Of the 324 empirical articles in this review, 22 used advanced text analytics (i.e., supervised or unsupervised machine learning). Of these, four were directly relevant to employment: Banks et al. (2019); Campion et al. (2016); Harrison et al. (2019a, 2019b); and Sajjadiani et al. (2019). All of these studies illustrated benefits with massive appeal to organizations and scholars alike. For example, they showed how these approaches and novel uses of available data can help inform and improve recruiting messages. These studies also show how the field has a substantial opportunity to use the more sophisticated CATA software that applies NLP to realize the benefits of CATA for employment research and decisions. The more sophisticated software programs have only recently become widely available (and affordable). This, plus the increasing recognition of the value of CATA, will likely result in an explosion of research using these methods in the next decade.

Takeaway #6

Due to the unobtrusiveness of some text data collection, we suspect that we may be able to significantly reduce opportunities and instances of faking in employment assessments. That is, prompting open-ended responses with little indication of what is being measured may provide a distinct advantage over the sometimes transparent Likert-type scales. Narrative information may be more difficult to fake because candidates only provide information based on a prompt and the computer scores it (e.g., measuring personality based on descriptions of past work behavior) as opposed to the candidates scoring themselves (e.g., self-report personality tests where candidates can decipher which responses give the highest scores). For example, Waung et al. (2017) used CATA to examine the types of impression management tactics candidates employed across resumes and cover letters by examining impression management-related words. As resumes and cover letters are hallmarks of the application process, the ability to quickly determine their level of impression management may be informative for employment decision makers.

Takeaway #7

Finally, limited research has been conducted on the topic of subgroup differences using CATA. We found only four studies that examined subgroup differences. In a study of the extent to which managers of different genders delegate, Akinola et al.

(2018) found that women were more likely to associate having to delegate with greater negative affect than men. In a study using text mining of applicant accomplishments, Campion et al. (2016) found that scoring essays with a supervised algorithm introduced no additional adverse impact than what was in the human ratings. They suggested that computer scoring against assessor scores should only produce adverse impact if there is already adverse impact in the assessor scores since the computer is simply modeling the assessors because text mined variables are only retained if they predict assessor scores. However, this should be the focus of future research because CATA variables may still capture ancillary variance associated with subgroup differences.

In another example, Kanze, Huang, Conley, and Higgins (2018) examined why male entrepreneurs raise more funding and found that investors ask female entrepreneurs more prevention-focused questions and ask males entrepreneurs more promotion-oriented questions. This type of study might be used to determine whether and why recruiters ask female candidates different questions than male candidates. Finally, Madera et al. (2009) measured characteristics of candidates (communal versus agentic traits) in letters of recommendation using LIWC and their own dictionary. The purpose of the study was to examine subgroup differences and they found that female candidates were described in more communal terms while male candidates were described in more agentic terms. Nevertheless, measuring communal and agency traits might be useful for hiring employees, with each more or less relevant to different job requirements. For example, communal traits might predict teamwork or citizenship performance, while agentic might predict task performance. However, the gender differences they found suggests that letter of references, if used, may create adverse impact if selection decisions emphasize agency.

There are a few final considerations regarding CATA and adverse impact that require our attention. Based on our understanding of current professional practice, as well as presentations at recent professional conferences (e.g., Walmsley, 2019), adverse impact is often observed when machine learning (including CATA) is applied to selection information (such as applications). Yet, there are several possible solutions to help address this:

1. When building models (e.g., retaining text mined variables) based on how well they predict criteria, the models should not show subgroup differences if the criteria do not show such differences, as was noted in Campion et al. (2016), although that should be confirmed by future research. However, this approach may not be helpful in many contexts because subgroup differences exist in many job-related KSAOs and performance in many jobs, so the criteria used in the modeling will show differences. In those situations, the researcher should instead evaluate whether using CATA creates any *additional* adverse impact beyond what already exists in the criteria.
2. Deny the computer from mining any information that might be illegal (e.g., race, gender, and age information). Of course, this includes the consideration of protected categories directly (discrimination), but it also includes indirect consideration from systematic error due to contamination by non-job-related

variance. An example of the latter that might be of special concern to CATA is the influence of writing skill when that is not job related. This might also include information that is associated with these protected categories (e.g., name of school attended). Only allowing the computer to consider information that is demonstrably job-related should avoid this concern (e.g., Campion et al., 2016). Note that machine learning (CATA included) is sometimes used to discover job-related variables, so the modeling cannot be based only on known job-related variables. In those cases, the researchers must ensure that the criterion used to retain the variables is job related and bias-free, as noted above, which will probably be based on content validity and other rational judgments or rely on one of the empirical approaches below.

3. Correlate the variables extracted by the computer with race and gender, and eliminate those variables that show differences. While useful, this approach has potential validity costs. As commonly observed in employment testing, the items with the largest race differences are often the most valid items, so eliminating them reduces validity (Ployhart & Holtz, 2008).
4. Program the algorithm to not have subgroup differences. This may potentially be tantamount to within-group norming, which is illegal based on the Civil Rights Act of 1991. In other words, adjusting test scores to make them equal across protected subgroups is illegal because it explicitly considers subgroup membership.

OBSERVATIONS ON CATA FROM OTHER DISCIPLINES

While management scholars are not necessarily CATA novices, our collective use of more advanced techniques remains relatively nascent. As such, we briefly turn to related disciplines to identify any lessons that could prove useful for management CATA researchers. The vast literature on machine learning was not directly relevant because it did not focus on human attributes and usually not on text data; however, a number of potentially insightful observations emerged.

First, machine learning research outside management often includes quantitative and text variables and management scholars should therefore not limit themselves to text data and CATA applications of machine learning. Quantitative variables can be included in the same computer algorithms as the text-mined variables. Most often, the text-mined variables are converted into quantitative variables and included in the statistical model (like in a regression). In the case of assessment for staffing, other data might include application material such as years of work experience, years of education, grades, or even test scores or other assessment information.

Second, the focus in the other literatures on prediction of outcomes over construct validity reveals that there are many other statistical models that might improve prediction. The management literature is unnecessarily narrow in its near sole reliance on ordinal least squares regression. Examples of other regression models available include logistic regression (for dichotomous criteria), cox regression (for hazard models), autoregressive integrated moving-average model (for time series), discriminant analysis (for distinguishing between groups), Poisson regression (for low probability events), and Gamma regression (for strictly positive but low data ranges). There are plenty of alternatives to regression in machine learning, as well, including cluster analyses (k-means), classification models (decision trees, support vector machines, nearest neighbor, etc.), neural networks, and Bayesian networks. Some of these techniques have been used in the management literature, but they are not as common. They may have advantages for improving the prediction of outcomes like job performance or turnover in HR management.

Third, the range of outcomes predicted in those other literatures might give ideas as to the broader applicability of these techniques. For example, research predicting customer behaviors from customer reviews (Bilro, Loureiro, & Guerreiro, 2019) may have insights for how to build and maintain employee loyalty and strengthen their brand identification (e.g., organizational reputation). Interestingly, a study of Amazon customer reviews showed that sentiment of reviews can be contagious and influence perceptions of how helpful a review is (Felbermayr & Nanopoulos, 2016). Although emotional contagion has been demonstrated in employment (e.g., Totterdell, 2000), few studies have assessed the impact of allowing employees to share views on various company topics (e.g., new policies). Furthermore, researchers in management tend to only look at job performance as a continuum, such as on a 5-point performance scale. Those in other literatures commonly predict specific categories (e.g., the likelihood someone will be rated a 5). Likewise, they would be more likely to try to predict subdimensions of job performance as opposed to just the overall composite like we do in management. Teasing apart the criterion in these ways can often lead to useful insights and predictions, as long as we properly attend to issues of capitalization on chance and spurious findings. The concept of "data mining" and its search for any and all possible relationships in the data may feel like a retreat to the "dust-bowl empiricism" days in management research, but proper statistical treatment and more focus on construct validity should attenuate those concerns.

Fourth, researchers in other fields will often make adjustments to the distributions of the data, including trimming and imputing data, and using assumed distributions. Management researchers tend to view the actual data collected as the best representation of the true phenomenon to be modeled or predicted, but these are usually convenience samples with known statistical limitations (e.g., range restriction, missing data, skew). As with relying on meta-analytic estimates of validity as opposed to performing a local validation where statistical limitations reduce the chances of finding validity, sometimes it is best to create a computer model based on the likely distribution rather than relying solely on available distributions with known problems. This is especially the case in the early stages of the research when the focus is on whether something "could" work as opposed to "will" work. This is much like conducting a lab study on a phenomenon to demonstrate its potential existence, and then following up with a field study to evaluate generalizability.

Data scientists will also use mathematical techniques to help solve problems with data distributions. As a specific example, the second author was asked to review a machine learning project for hiring at a large organization. The model used application information to predict recruiter decisions and subsequent job performance. The problem the company faced was that, due to minimum qualifications, there is almost no variance in such variables and certain candidate characteristics would be almost completely missing from data. For example, the possession of a college degree is likely a minimum qualification for hiring software engineers and therefore no one is without such a qualification, making it difficult to use this type of construct for prediction. This company addressed the problem by using "inverse propensity weighting" wherein the model weights were adjusted to over-count the small sample of those hired without degrees.

Finally, the broader machine learning literature shows that the sources, styles, and amount of data are seemingly endless. As noted previously, these techniques have broad applicability and include data we have historically ignored as being useful for the purposes of traditional HRM topics (e.g., tweets, online reviews). Furthermore, the capacity of operating systems to process and manage huge amounts of data should reduce our apprehension about these data sources. For example, one study we found while examining research in other literatures text mined more than 1.7 million tweets to analyze the sentiment of particular brands (e.g., Comcast) (Liu, Burns, & Hou, 2017). While CATA research in management has used a wide range of text data sources, we are hopeful for future studies to take advantage of additional untapped data sources (such as tweets).

CHALLENGES IN THE USE OF TEXT ANALYSIS IN EMPLOYMENT

Broadly speaking, CATA is challenging for those inside and outside of management. To assist researchers and organizations that are considering the use of CATA to improve scholarship and employment processes, we discuss some of the challenges at each stage below. We end this chapter with a list of recommendations on how to address or prevent such challenges.

1. *Identifying viable applications*. Because there are many potential CATA uses, this decision requires a number of considerations. For example, if text data do not currently exist (e.g., on applications or employee records), then determining whether and how it will be collected is primary. Is the textual data collected likely to contain information on job-related constructs? If not, could the data collected be modified? For example, in the context of hiring, could the application be modified to ensure it collects relevant information (e.g., details on past job tasks, courses taken in school, evidence of claimed special skills, etc.), could questions be asked that require narrative descriptions (e.g., past work behavior or accomplishments), and can the application process be modified to require completing a standardized application as opposed to accepting resumes as sufficient to apply? Moreover, whether the data collection is proctored or unproctored has implications in a hiring context because candidates may receive help with their answers or their writing when unproctored.

   In addition to ensuring the quality of the content of the textual data, there are several statistical challenges such as ensuring that there is enough variability in the data to be useful for selection. Young candidates might not have enough work experience to provide sufficient information on past work behavior or accomplishments to be analyzed. Moreover, past work experience or education may not be highly relevant. Relatedly, researchers and organizations must collect a large enough sample to create the CATA model. In one example, Campion et al. (2016) examined the sample size question and suggested 500 responses as a minimum rule-of-thumb with their data. The size of the model additionally informs the appropriate sample size because the more parameters, the greater the sample needed for stable estimates.

2. *Deciding on which software to use*. As stated above, researchers should seek to use more sophisticated software beyond the data-management packages and the dictionary-based systems. The authors are inclined to use commercial software because it is more fully developed, user friendly, documented, and technically supported. Furthermore, most commercial software does not require programing skills, which is a challenge for management scholars who are trained in statistics and theory rather than software coding.
3. *Learning the software*. Assuming the researcher uses more sophisticated software packages, there will be a meaningful but not prohibitive investment in time to learn, even if commercial software is used. How much time will depend on many factors (e.g., computer skill, availability of a trainer, complexity of data, etc.). We estimate it will likely take several days to a week for a researcher to learn the basics, and then several weeks or months of additional time learning "on-the-job" as issues emerge while applying the software.
4. *Training the software*. This again assumes the use of more sophisticated software and can be a time-consuming task. Training involves working with the computer model incrementally to improve it and can include revising the concepts extracted by the computer or the aggregate categories of concepts. Training may take the form of combining, deleting, relabeling, identifying synonyms, and other adjustments. A primary factor influencing the difficulty of training is the range of possible responses because it influences the number of CATA categories needed. For example, a CATA model for a personal statement by candidates will require more categories than a specific question on past leadership experience. Training may also involve the validation step by selecting the text-mined variables to retain based on their correlation with some criterion. A logical question regarding training is: how much training is enough training? In some previous research (Campion et al., 2016), the goal was to train the computer model until the correlation with a human assessor was the same as the correlation between assessors. However, another goal might be to continue training until some asymptotic level is reached. Whether based on statistical criteria or on the theoretical or practical appeal of the model, the burden to defend the final model is on the researcher.
5. *Validating the model*. This is perhaps one of the most important and yet challenging steps in using CATA for employment research and decisions. Due to legal requirements and social expectations around employment decision

making, management scholars and organizations are held to higher standards when using advanced techniques and must be able to support their validity (including face validity). There are myriad decisions scholars must make in this realm. For example, if content validation is to be used, what will be the approach? Is a job analysis available? Are SMEs available? Will linkage analyses be conducted? Would the context meet the requirements for content validation in the various legal and professional validation guidelines? While we have noted that a useful approach to training is to do so against a criterion, it may be that researchers do not have a criterion to text mine against. In this case, it is particularly crucial that the researcher clearly describe the decision to maintain the model (e.g., a theory supports the model). When a criterion is available, challenges include, but are not limited to, restriction of range in the predictor, unreliability of the criterion, and statistical power.

6. *Updating the model to accommodate legal and social changes*. CATA models may require revision due to changes in employment laws (e.g., Illinois Artificial Intelligence Video Interview Act), but also due to social developments. Words can change or alter meaning over time. For example, the word "literally" is now well-understood to metaphorically mean "figuratively," and while the current generation was not the first to use it this way, they were the ones to make it popular (Merriam-Webster, n.d.; also see "Google" as a dictionary addition in 2006). Colloquial expressions are especially likely to change over time. Researchers in organizations are responsible for updating their models to meet legal requirements and accommodate relevant social changes.
7. *Working out operational details*. As with any project, the "devil is in the details." Creating a CATA model for operational use in hiring and putting it into practice is no exception and may be even more complicated due to the need for information technology support and integration with the company's existing computer systems. Issues include, but are not limited to, researcher support, programmer support, hardware requirements, flow of data and time, data cleaning, scoring, cut scores, data maintenance, security issues, and other integration issues.
8. *Communicating to candidates*. Although the use of CATA to help score textual data as part of the hiring process does not necessarily create additional needs to communicate with candidates, the usual communications may have to be adjusted. These will likely include communications posted on the website or in other recruiting documents as to the nature of the assessment used in the hiring process, any preparation advice, feedback on scores, retesting policies, responses to Freedom of Information Act (FOIA) requests for government employers, and so on. However, some candidates may still have mistrust in computer scoring that will influence their reaction (Gonzalez, Capman, Oswald, Theys, & Tomczak, 2019). Perhaps communicating that the machine scoring is highly reliable, bias-free, and involves no human subjectivity would mitigate their concerns as fairness communications have been shown to improve reactions in other hiring contexts (e.g., Truxillo, Bauer, Campion, & Paronto, 2002).

RECOMMENDATIONS FOR USING CATA

Many recommendations have already been made explicitly or implicitly throughout the article for using CATA. The current section summarizes our recommendations regarding the approaches to use in research on employment decision making and HRM more generally.

1. In terms of a bottom-line, we recommend that CATA be used as an approach to measuring KSAOs for employment decisions. Our review also suggests that CATA can and has been used to measure a range of constructs. Although relatively few have measured KSAs (other than Os) directly related to employment decisions, the range of constructs measured in the literature suggests that KSAs could viably be measured. In fact, they might be measured more easily because they are more definable, objective, and verifiable than many psychological constructs measured in the literature.
2. We recommend using more sophisticated methods than the CATA methods used in most of the management literature to date. They are generally very simplistic, such as merely facilitating manual content analyses or employing an automated word search in a corpus of words that represents a particular construct (e.g., entrepreneurial orientation). More sophisticated and better approaches are now available and are capable of automating the relationships among words rather than simply counting individual words.
3. We recommend using approaches to CATA that evaluate strings of words and relationships among words like NLP, as opposed to the more common and simple single word and phrase-based approaches. The true benefits of CATA will only be realized with these more sophisticated and automated programs.
4. We recommend using approaches that allow training. Approaches vary in terms of whether and how easily they can be trained. The literature will often use the terms "supervised versus unsupervised" to make this distinction. Organizations using CATA for employment will want to use approaches that allow training because it can greatly improve interpretation and prediction. Another value of training is that it helps ensure that the researcher understands what the computer is scoring instead of blindly accepting the "black box." This is similar to how researchers must learn how to calculate statistics by hand in graduate school, even though they will use computers in the future. Parenthetically, the layperson interpretation of "machine learning" is that the computer teaches itself. However, in the current state of development of the field, this refers more to the fact that the computer can fit a model to the data (like regression has done for years), rather than some continuous self-teaching from moment to moment. Researchers still have to train the computer in many ways even with today's modern software.
5. This all said, we also recommend not ignoring the use of word dictionaries. They offer at least two advantages. First, many dictionaries have already been developed and validated on a range of constructs, and they are inexpensive to purchase (e.g., LIWC, DICTION). Second, it is fairly easy to develop one's own dictionary on constructs of interest. Although data dictionaries are

perhaps the most simplistic approach to text mining, they can be developed in advance based on theory and do not rely on a corpus to text mine necessarily as the first step like the more sophisticated approaches.

6. Relatedly, we recommend considering the various ways to strategically select or create corpuses (corpora) for developing a CATA model. For example, documents could be selected that are likely to be enriched with words relevant to constructs of interest in order to identify words for models (e.g., documents describing leadership might be used to identify leadership-related word descriptions). Similarly, illustrative text samples created by SMEs can be text mined to measure constructs better than using actual examples from subjects in some instances. Creating an algorithm for scoring constructs does not always start with actual examples from the future intended corpora.
7. We also recommend considering sentiment analysis as it might be appropriate. It could conceivably be used anytime it is necessary to distinguish the positive or negative tenor of comments. This has obvious applications when analyzing survey responses where the tone of the comments is important along with the content. It could potentially help resolve one of the most central problems in performance evaluation: leniency and skew in the ratings due to the unwillingness of managers to give candid feedback. Narrative comments are somewhat less susceptible to this issue. These comments may be more candid because, unlike the ratings, they are not usually directly linked to decisions like compensation (Speer, 2018). These comments are also less susceptible because, even though they will be consistent with the ratings, they can reveal the strength of the job performance through "faint praise." Sentiment-based CATA might be able to measure these nuances more objectively. Similarly, employers want to hire candidates with a "positive attitude," but this is vulnerable to faking in some traditional assessments (e.g., interviews). Using sentiment analysis to assess attitudes from written materials may make response bias more difficult because it is less direct (less obvious to candidates) and it will be more difficult to determine which words are scored.
8. Finally, we recommend considering some of the insights from the broader machine learning literature. This includes using both quantitative and text variables in the models, using other statistical models that might improve prediction, expanding the range of outcomes predicted, making adjustments to data distributions as appropriate, and broadening the sources of data considered.

NOTES

1. For example, it is been recognized in some capacity for the last 7 years in the Society for Industrial and Organizational Psychologists annual work trends surveys (https://www.siop.org/Business-Resources/Top-10-Workplace-Trends).

2. We searched the following journals: *Academy of Management Annals, Academy of Management Journal, Academy of Management Proceedings, Administrative Science Quarterly, Annual Review of Organizational Psychology and Behavior, Entrepreneurship Theory and Practice, Journal of Applied Psychology, Journal of Business Venturing, Journal of Business and Psychology, Journal of International Business Studies, Journal of Management, Journal of Organizational Behavior, Organizational Behavior and Human Decision Processes,*

Organizational Research Methods, Organization Science, Personnel Psychology, and Strategic Management Journal.

3. An exception might be when an organization uses an automated writing assessment as part of its hiring process. For a recent review, see Shermis and Burstein (2013).

REFERENCES

Akinola, M., Martin, A. E., & Phillips, K. W. (2018). To delegate or not to delegate: Gender differences in affective associations and behavioral responses to delegation. *Academy of Management Journal*, *61*(4), 1467–1491.

Anglin, A. H., Wolfe, M. T., Short, J. C., McKenny, A. F., & Pidduck, R. J. (2018). Narcissistic rhetoric and crowdfunding performance: A social role theory perspective. *Journal of Business Venturing*, *33*(6), 780–812.

Antons, D., Joshi, A. M., & Salge, T. O. (2018). Content, contribution, and knowledge consumption: Uncovering hidden topic structure and rhetorical signals in scientific texts. *Journal of Management*, *45*(7), 3035–3076.

Banks, G. C., Woznyj, H. M., Wesslen, R. S., Frear, K. A., Berka, G., Heggestad, E. D., & Gordon, H. L. (2019). Strategic recruitment across borders: An investigation of multinational enterprises. *Journal of Management*, *45*(2), 476–509.

Barley, S. R., Meyerson, D. E., & Grodal, S. (2011). E-mail as a source and symbol of stress. *Organization Science*, *22*(4), 887–906.

Barlow, M. A., Verhaal, J. C., & Hoskins, J. D. (2018). Guilty by association: Product-level category stigma and audience expectations in the US craft beer industry. *Journal of Management*, *44*(7), 2934–2960.

Basu, S., & Savani, K. (2017). Choosing one at a time? Presenting options simultaneously helps people make more optimal decisions than presenting options sequentially. *Organizational Behavior and Human Decision Processes*, *139*, 76–91.

Beckman, C. M., & Stanko, T. L. (2019). It takes three: Relational boundary work, resilience, and commitment among navy couples. *Academy of Management Journal*. (Advanced online publication.)

Bednar, M. K. (2012). Watchdog or lapdog? A behavioral view of the media as a corporate governance mechanism. *Academy of Management Journal*, *55*(1), 131–150.

Bilro, R. G., Loureiro, S. M. C., & Guerreiro, J. (2019). Exploring online customer engagement with hospitality products and its relationship with involvement, emotional states, experience and brand advocacy. *Journal of Hospitality Marketing & Management*, *28*(2), 147–171.

Bligh, M. C., Kohles, J. C., & Meindl, J. R. (2004). Charisma under crisis: Presidential leadership, rhetoric, and media responses before and after the September 11th terrorist attacks. *The Leadership Quarterly*, *15*(2), 211–239.

Boling, J. R., Pieper, T. M., & Covin, J. G. (2015). CEO tenure and entrepreneurial orientation within family and nonfamily firms. *Entrepreneurship Theory & Practice*, *706*, 864–1614.

Buyl, T., Boone, C., & Wade, J. B. (2019). CEO narcissism, risk-taking, and resilience: An empirical analysis in US commercial banks. *Journal of Management*, *45*(4), 1372–1400.

Campion, M. C., Campion, M. A., Campion, E. D., & Reider, M. H. (2016). Initial investigation into computer scoring of candidate essays for personnel selection. *Journal of Applied Psychology*, *101*(7), 958–975.

Carley, K. M. (1997). Extracting team mental models through textual analysis. *Journal of Organizational Behavior*, *18*, 533–558.

Chatman, J. A., Caldwell, D. F., O'Reilly, C. A., & Doerr, B. (2014). Parsing organizational culture: How the norm for adaptability influences the relationship between culture consensus and financial performance in high-technology firms. *Journal of Organizational Behavior*, *35*(6), 785–808.

Chen, X., & Latham, G. P. (2014). The effect of priming learning vs. performance goals on a complex task. *Organizational Behavior and Human Decision Processes*, *125*(2), 88–97.

Chuang, A., Hsu, R. S., Wang, A. C., & Judge, T. A. (2015). Does West "fit" with East? In search of a Chinese model of person–environment fit. *Academy of Management Journal*, *58*(2), 480–510.

Clark, D. R., Li, D., & Shepherd, D. A. (2018). Country familiarity in the initial stage of foreign market selection. *Journal of International Business Studies*, *49*, 442–472.

Connelly, C. E., Zweig, D., Webster, J., & Trougakos, J. P. (2012). Knowledge hiding in organizations. *Journal of Organizational Behavior*, *33*(1), 64–88.

Corner, P. D., & Ho, M. (2010). How opportunities develop in social entrepreneurship. *Entrepreneurship Theory and Practice*, *34*(4), 635–659.

Crayne, M. P., & Hunter, S. T. (2018). Historiometry in organizational science: Renewed attention for an established research method. *Organizational Research Methods*, *21*(1), 6–29.

Creed, W. D., DeJordy, R., & Lok, J. (2010). Being the change: Resolving institutional contradiction through identity work. *Academy of Management Journal*, *53*(6), 1336–

Crilly, D. (2017). Time and space in strategy discourse: Implications for intertemporal choice. *Strategic Management Journal*, *3*, 2370–2389.

Detert, J. R., & Treviño, L. K. (2010). Speaking up to higher-ups: How supervisors and skip-level leaders influence employee voice. *Organization Science*, *21*(1), 249–270.

DICTION. (2019). Retrieved from https://www.dictionsoftware.com/diction-overview/

Edmondson, A. C., & McManus, S. E. (2007). Methodological fit in management field research. *Academy of Management Review*, *32*(4), 1246–1264.

Engelen, A., Neumann, C., & Schwens, C. (2015). "Of course I can": The effect of CEO overconfidence on entrepreneurially oriented firms. *Entrepreneurship Theory and Practice*, *39*(5), 1137–1160.

Engelen, A., Neumann, C., & Schmidt, S. (2016). Should entrepreneurially oriented firms have narcissistic CEOs?. *Journal of Management*, *42*(3), 698–721.

Fanelli, A., Misangyi, V. F., & Tosi, H. L. (2009). In charisma we trust: The effects of CEO charismatic visions on securities analysts. *Organization Science*, *20*(6), 1011–1033.

Felbermayr, A., & Nanopoulos, A. (2016). The role of emotions for the perceived usefulness in online customer reviews. *Journal of Interactive Marketing*, *36*, 60–76.

Felps, W., Mitchell, T. R., Hekman, D. R., Lee, T. W., Holtom, B. C., & Harman, W. S. (2009). Turnover contagion: How coworkers' job embeddedness and job search behaviors influence quitting. *Academy of Management Journal*, *52*(3), 545–561.

Field, A., Miles, J., & Field, Z. (2012). *Discovering Statistics Using R*. London: Sage.

Follmer, E. H., Talbot, D. L., Kristof-Brown, A. L., Astrove, S. L., & Billsberry, J. (2018). Resolution, relief, and resignation: A qualitative study of responses to misfit at work. *Academy of Management Journal*, *61*(2), 440–465.

Gephart, R. P. (2004). Qualitative research in the Academy of Management Journal. *Academy of Management Journal*, *47*(4), 454–462.

Gibson, C. B., & Gibbs, J. L. (2006). Unpacking the concept of virtuality: The effects of geographic dispersion, electronic dependence, dynamic structure, and national diversity on team innovation. *Administrative Science Quarterly*, *51*(3), 451–495.

Gibson, C. B., Gibbs, J. L., Stanko, T. L., Tesluk, P., Cohen, S. G. (2011). Including the "I" in Virtuality and Modern Job Design: Extending the Job Characteristics Model to Include the Moderating Effect of Individual Experiences of Electronic Dependence and Copresence. *Organization Science*, *22*, 1481–1499.

Gioia, D. A., Price, K. N., Hamilton, A. L., & Thomas, J. B. (2010). Forging an identity: An insider-outsider study of processes involved in the formation of organizational identity. *Administrative Science Quarterly*, *55*(1), 1–46.

Glaser, B., & Strauss, A. (1967). Grounded theory: The discovery of grounded theory. *Sociology the Journal of the British Sociological Association*, *12*(1), 27–49.

Gonzalez, M. F., Capman, J. F., Oswald, F. L., Theys, E. R., & Tomczak, D. L. (2019). Where's the I-O?" Artificial intelligence and machine learning in talent management systems. *Personnel Assessment and Decisions*, *5*(3), 33–44.

Grühn, B., Strese, S., Flatten, T. C., Jaeger, N. A., & Brettel, M. (2017). Temporal change patterns of entrepreneurial orientation: A longitudinal investigation of CEO successions. *Entrepreneurship Theory and Practice*, *41*(4), 591–619.

Hackman, J. R., & Oldham, G. R. (1975). Development of the job diagnostic survey. *Journal of Applied Psychology*, *60*(2), 159–170.

Harrison, J. S., Thurgood, G. R., Boivie, S., & Pfarrer, M. D. (2019a). Measuring CEO personality: Developing, validating, and testing a linguistic tool. *Strategic Management Journal*, *40*(8), 1316–1330.

Harrison, J. S., Thurgood, G. R., Boivie, S., & Pfarrer, M. D. (2019b). Perception is reality: How ceos' observed personality influences market perceptions of firm risk and shareholder returns. *Academy of Management Journal*, doi:https://doi.org/10.5465/amj.2018.0626.

Hartwell, C. J., & Campion, M. A. (2019). Getting social in selection: How social networking website content is perceived and used in hiring. *International Journal of Selection and Assessment*.

Hirsh, J. B., & Peterson, J. B. (2009). Personality and language use in self-narratives. *Journal of Research in Personality*, *43*, 524–527.

Janasik, N., Honkela, T., & Bruun, H. (2009). Text mining in qualitative research: Application of an unsupervised learning method. *Organizational Research Methods*, *12*(3), 436–460.

Josefy, M., Dean, T. J., Albert, L. S., & Fitza, M. A. (2017). The role of community in crowdfunding success: Evidence on cultural attributes in funding campaigns to "Save the local theater". *Entrepreneurship Theory and Practice*, *41*(2), 161–182.

Kabanoff, B., & Holt, J. (1996). Changes in the espoused values of Australian organizations 1986—1990. *Journal of Organizational Behavior*, *17*(3), 201–219.

Kabanoff, B., Waldersee, R., & Cohen, M. (1995). Espoused values and organizational change themes. *Academy of Management Journal*, *38*(4), 1075–1104.

Kang, J., & Kim, A. Y. H. (2017). The relationships between CEO media appearances and compensation. *Organization Science*, *28*(3), 379–394.

Kanze, D., Huang, L., Conley, M. A., & Higgins, E. T. (2018). We ask men to win and women not to lose: Closing the gender gap in startup funding. *Academy of Management Journal*, *61*(2), 586–614.

Keil, T., Maula, M., & Syrigos, E. (2017). CEO entrepreneurial orientation, entrenchment, and firm value creation. *Entrepreneurship Theory and Practice*, *41*(4), 475–504.

Kellogg, K. C., Orlikowski, W. J., & Yates, J. (2006). Life on the trading zone: Structuring coordination across boundaries in postbureaucratic organizations. *Organization Science*, *17*(1), 22–44.

Kobayashi, V. B., Mol, S. T., Berkers, H. A., Kismihok, G., & Den Hartog, D. N. (2018a). Text classification for organizational researchers: A tutorial. *Organizational Research Methods*, *21*(3), 766–799.

Kobayashi, V. B., Mol, S. T., Berkers, H. A., Kismihok, G., & Den Hartog, D. N. (2018b). Text mining in organizational research. *Organizational Research Methods*, *21*(3), 733–765.

Kreiner, G. E., Hollensbe, E. C., & Sheep, M. L. (2009). Balancing borders and bridges: Negotiating the work-home interface via boundary work tactics. *Academy of Management Journal*, *52*(4), 704–730.

Lanaj, K., Foulk, T. A., & Erez, A. (2019). Energizing leaders via self-reflection: A within-person field experiment. *Journal of Applied Psychology*, *104*(1), 1–18.

Laureiro-Martinex, D., & Brusoni, S. (2017). Cognitive flexibility and adaptive decision-making: Evidence from a laboratory study of expert decision makers. *Strategic Management Journal*, *39*, 1031–1058.

Liang, L. H., Lian, H., Brown, D. J., Ferris, D. L., Hanig, S., & Keeping, L. M. (2016). Why are abusive supervisors abusive? A dual-system self-control model. *Academy of Management Journal*, *59*(4), 1385–1406.

Liu, X., Burns, A. C., & Hou, Y. (2017). An Investigation of Brand-Related User-Generated Content on Twitter. *Journal of Advertising*, *46*(2), 236–247.

Locke, K. (2001). *Grounded theory in management research*. Thousand Oaks, CA: Sage Publications.

Love, E. G., Lim, J., & Bednar, M. K. (2017). The face of the firm: The influence of CEOs on corporate reputation. *Academy of Management Journal*, *60*(4), 1462–1481.

Luciano, M. M., Mathieu, J. E., Park, S., & Tannenbaum, S. I. (2018). A fitting approach to construct and measurement alignment: The role of big data in advancing dynamic theories. *Organizational Research Methods*, *21*(3), 592–632.

Madera, J. M., Hebl, M. R., & Martin, R. C. (2009). Gender and letters of recommendation for academia: Agentic and communal differences. *Journal of Applied Psychology*, *94*(6), 1591.

Malhotra, S., Reus, T. H., Zhu, P., & Roelofsen, E. M. (2018). The acquisitive nature of extraverted CEOs. *Administrative Science Quarterly*, *63*(2), 370–408.

Martin, S. R. (2016). Stories about values and valuable stories: A field experiment of the power of narratives to shape newcomers' actions. *Academy of Management Journal*, *59*(5), 1707–1724.

McAlearney, A. S. (2006). Leadership development in healthcare: a qualitative study. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, *27*(7), 967–982.

McKenny, A. F., Aguinis, H., Short, J. C., & Anglin, A. H. (2018). What doesn't get measured does exist: Improving the accuracy of computer-aided text analysis. *Journal of Management*, *44*(7), 2909–2933.

McKenny, A. F., Short, J. C., & Payne, G. T. (2013). Using computer-aided text analysis to elevate constructs: An illustration using psychological capital. *Organizational Research Methods*, *16*(1), 152–184.

Merriam-Webster (n.d.). Retrieved from https://www.merriam-webster.com/words-at-play/misuse-of-literally

Moore, C., Lee, S. Y., Kim, K., & Cable, D. M. (2017). The advantage of being oneself: The role of applicant self-verification in organizational hiring decisions. *Journal of Applied Psychology*, *102*(11), 1493–1513.

Moss, T. W., Neubaum, D. O., & Meyskens, M. (2015). The effect of virtuous and entrepreneurial orientations on microfinance lending and repayment: A signaling theory perspective. *Entrepreneurship Theory and Practice*, *39*(1), 27–52.

Moss, T. W., Renko, M., Block, E., & Meyskens, M. (2018). Funding the story of hybrid ventures: Crowdfunder lending preferences and linguistic hybridity. *Journal of Business Venturing*, *33*(5), 643–659.

Nadkarni, S., & Narayanan, V. K. (2005). Validity of the structural properties of text-based causal maps: An empirical assessment. *Organizational Research Methods*, *8*(1), 9–40.

Newman, D. A., Harrison, D. A., Carpenter, N. C., & Rariden, S. M. (2016). Construct mixology: Forming new management constructs by combining old ones. *Academy of Management Annals*, *10*(1), 943–995.

Osborne, J. D., Stubbart, C. I., & Ramaprasad, A. (2001). Strategic groups and competitive enactment: A study of dynamic relationships between mental models and performance. *Strategic Management Journal*, *22*, 435–454.

Owens, B. P., & Hekman, D. R. (2012). Modeling how to grow: An inductive examination of humble leader behaviors, contingencies, and outcomes. *Academy of Management Journal*, *55*(4), 787–818.

Pandey, S., & Pandey, S. K. (2017). Applying natural language processing capabilities in computerized textual analysis to measure organizational culture. *Organizational Research Methods*, *22*(3), 765–797.

Peltokorpi, V., & Vaara, E. (2014). Knowledge transfer in multinational corporations: Productive and counterproductive effects of language-sensitive recruitment. *Journal of International Business Studies*, *45*(5), 600–622.

Pennebaker, J. W., Chung, C. K., Ireland, M., Gonzales, A., & Booth, R. J. (2007). The development and psychometric properties of LIWC2007. Retrieved from http://www.liwc.net/LIWC2007LanguageManual.pdf

Peters, L. H., & O'Connor, E. J. (1980). Situational constraints and work outcomes: The influences of a frequently overlooked construct. *Academy of management review*, *5*(3), 391–397.

Peterson, N. G., Mumford, M. D., Borman, W. C., Jeanneret, P. R., Fleishman, E. A., Levin, K. Y., … Dye, D. (2001). Understanding work using the occupational information network (O*NET): Implications for practice and research. *Personnel Psychology*, *54*, 451–492.

Petriglieri, J. L. (2015). Co-creating relationship repair: Pathways to reconstructing destabilized organizational identification. *Administrative Science Quarterly*, *60*(3), 518–557.

Ployhart, R. E., & Holtz, B. C. (2008). The diversity–validity dilemma: Strategies for reducing racioethnic and sex subgroup differences and adverse impact in selection. *Personnel Psychology*, *61*(1), 153–172.

Reay, T., Goodrick, E., Waldorff, S. B., & Casebeer, A. (2017). Getting leopards to change their spots: Co-creating a new professional role identity. *Academy of Management Journal*, *60*(3), 1043–1070.

Reilly, R. R., & Chao, G. T. (1982). Validity and fairness of some alternative employee selection procedures. *Personnel Psychology*, *35*(1), 1–62.

Reinecke, J., & Ansari, S. (2015). When times collide: Temporal brokerage at the intersection of markets and developments. *Academy of Management Journal*, *58*(2), 618–648.

Ridge, J. W., & Ingram, A. (2017). Modesty in the top management team: Investor reaction and performance implications. *Journal of Management*, *43*(4), 1283–1306.

Rothausen, T. J., Henderson, K. E., Arnold, J. K., & Malshe, A. (2017). Should I stay or should I go? Identity and well-being in sensemaking about retention and turnover. *Journal of Management*, *43*(7), 2357–2385.

Sajjadiani, S., Sojourner, A. J., Kammeyer-Mueller, J. D., & Mykerezi, E. (2019). Using machine learning to translate applicant work history into predictors of performance and turnover. *Journal of Applied Psychology*. (Advanced online publication).

SAS. (2019). Natural language processing: What it is and why it matters. Retrieved from https://www.sas.com/en_us/insights/analytics/what-is-natural-language-processing-nlp.html

Schmiedel, T., Müller, O., & vom Brocke, J. (2019). Topic modeling as a strategy of inquiry in organizational research: A tutorial with an application example on organizational culture. *Organizational Research Methods*, *22*(4), 941–968.

Schmidt, F., & Hunter, J. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124*, 262–274.

Shantz, A., & Latham, G. P. (2009). An exploratory field experiment of the effect of subconscious and conscious goals on employee performance. *Organizational Behavior and Human Decision Processes*, *109*(1), 9–17.

Shermis, M. D., & Burstein, J. (Eds.). (2013). *Handbook of automated essay evaluation: Current applications and new directions*. New York, NY: Routledge.

Short, J. C., Broberg, J. C., Cogliser, C. C., & Brigham, K. H. (2010). Construct validation using computer-aided text analysis (CATA) an illustration using entrepreneurial orientation. *Organizational Research Methods*, *13*(2), 320–347.

Short, J. C., McKenny, A. F., & Reid, S. W. (2018). More than words? Computer-aided text analysis in organizational behavior and psychology research. *Annual Review of Organizational Psychology and Organizational Behavior*, *5*, 415–435.

Shortt, H. L., & Warren, S. K. (2019). Grounded visual pattern analysis: Photographs in organizational field studies. *Organizational Research Methods*, *22*(2), 539–563.

Silge, J., & Robinson, D. (2017). *Text Mining with R: A Tidy Approach*. Sebastopol, CA: O'Reilly Media.

Silverman, D., & Marvasti, A. (2008). *Doing qualitative research: A comprehensive guide*. Thousand Oaks, CA: Sage.

SIOP Top 10 Workplace Trends. (n.d.). Retrieved from https://www.siop.org/Business-Resources/Top-10-Workplace-Trends

Slutskaya, N., Game, A. M., & Simpson, R. C. (2018). Better together: Examining the role of collaborative ethnographic documentary in organizational research. *Organizational Research Methods*, *21*(2), 341–365.

Smither, J. W., & Walker, A. G. (2004). Are the characteristics of narrative comments related to improvement in multirater feedback ratings over time?. *Journal of Applied Psychology*, *89*(3), 575.

Sonenshein, S. (2014). How organizations foster the creative use of resources. *Academy of Management Journal*, *57*(3), 814–848.

Speer, A. B. (2018). Quantifying with words: An investigation of the validity of narrative-derived performance scores. *Personnel Psychology*, *71*(3), 299–333.

Stanko, T. L., & Beckman, C. M. (2015). Watching you watching me: Boundary control and capturing attention in the context of ubiquitous technology use. *Academy of Management Journal*, *58*(3), 712–738.

Strauss, A. & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Newbury Park, NJ: Sage.

Titus, Jr, V., Parker, O., & Covin, J. (2019). Organizational aspirations and external venturing: the contingency of entrepreneurial orientation. *Entrepreneurship Theory and Practice*. doi:https://doi.org/10.1177/1042258719838473

Toegel, G., Kilduff, M., & Anand, N. (2013). Emotion helping by managers: An emergent understanding of discrepant role expectations and outcomes. *Academy of Management Journal*, *56*(2), 334–357.

Totterdell, P. (2000). Catching moods and hitting runs: Mood linkage and subjective performance in professional sport teams. *Journal of Applied Psychology*, *85*(6), 848–859

Truxillo, D. M., Bauer, T. N., Campion, M. A., & Paronto, M. E. (2002). Selection fairness information and applicant reactions: A longitudinal field study. *Journal of Applied Psychology*, *87*, 1020–1031.

Walmsley, P. T. (2019). Using machine learning and deep learning in hiring: Ethical, legal, and practical concerns. Panel discussion presented at the Society for Industrial and Organizational Psychology, National Harbor, Maryland.

Watson, A., Dada, O., Wright, O., & Perrigot, R. (2019). Entrepreneurial orientation rhetoric in *franchise organizations: The impact of national culture*. *Entrepreneurship Theory and Practice*, *43*(4), 751–772.

Waung, M., McAuslan, P., DiMambro, J. M., & Mięgoć, N. (2017). Impression management use in resumes and cover letters. *Journal of Business and Psychology*, *32*(6), 727–746.

Wilhelmy, A., Kleinmann, M., König, C. J., Melchers, K. G., & Truxillo, D. M. (2016). How and why do interviewers try to make impressions on applicants? A qualitative study. *Journal of Applied Psychology*, *101*(3), 313–332.

Wilson, K. S., DeRue, D. S., Matta, F. K., Howe, M., & Conlon, D. E. (2016). Personality similarity in negotiations: Testing the dyadic effects of similarity in interpersonal traits and the use of emotional displays on negotiation outcomes. *Journal of Applied Psychology*, *101*(10), 1405–1421.

Wolfe, M. T., & Shepherd, D. A. (2015). "Bouncing back" from a loss: Entrepreneurial orientation, emotions, and failure narratives. *Entrepreneurship Theory and Practice*, *39*(3), 675–700.

Zuzul, T. (2018). "Matter Battles:" Boundary Objects and the Failure of Collaboration in Two Smart Cities. *Academy of Management Journal*, *62*(3), 739–764.